nce Library

OCT 1 7 1945

Vol. 35, No. 5 October, 1945

# Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, ON LEAVE
FRANCIS W. IRWIN, ACTING EDITOR
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Southern California;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W. MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

## CONTENTS

*Errors of Visual Recognition and the Nature of the Trace:* D. O. HEBB AND ESME N. FOORD ..... 335

*Reasoning in Humans. III. The Mechanisms of Equivalent Stimuli and of Reasoning:* N. R. F. MAIER ..... 349

*The Effects of Electroshock Convulsions on Double Alternation Lever-Pressing in the White Rat:* E. MCGINNIES AND H. SCHLOSBERG ..... 361

*Tonal Range and Volume Level Preferences of Broadcast Listeners:* P. EISENBERG AND H. A. CHINN ..... 374

*A Sensitized Eyelid Reaction Related to the Conditioned Eyelid Response:* D. A. GRANT ..... 393

*Evidence for a Universal Shock Factor in Learning:* G. S. SNODDY ..... 403

*Reliability of Blinking Frequency Employed as a Measure of Readability:* M. A. TINKER ..... 418

*Artifacts in the Investigation of Sensitivity to Vibration:* L. D. GOODFELLOW 425

PUBLISHED BI-MONTHLY,
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879

PUBLICATIONS OF

# THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

## PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
*Princeton University*

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
*Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.*

## PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
*University of Minnesota*

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 665 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
*Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.*

## JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave,* FRANCIS W. IRWIN, *Acting Editor*
*University of Pennsylvania*

Contains original contributions of an experimental character. Appears bi-monthly (beginning 1944), one volume per year, each volume of six numbers containing about 520 pages.
*Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.*

## PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
*Brown University*

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
*Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.*

## PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
*University of North Carolina*

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 350 pages.
*Subscription: $6.00 per volume (Foreign, $6.30).*

## JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
*Harvard University*

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 400 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
*Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.*

## JOURNAL OF APPLIED PSYCHOLOGY

DONALD G. PATERSON, *Editor*
*University of Minnesota*

Covers the applications of psychology in business, industry, education, etc. Appears bi-monthly, February, April, June, August, October, and December.
*Subscription: $6.00. Single copies, $1.25.*

---

*Subscriptions, orders, and business communications should be sent to*

# THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Journal of
Experimental Psychology

VOL. 35, NO. 5 OCTOBER, 1945

# ERRORS OF VISUAL RECOGNITION AND THE NATURE OF THE TRACE

BY D. O. HEBB

*Yerkes Laboratories of Primate Biology*

AND

ESME N. FOORD

*Queen's University*

The Gestalt hypothesis of dynamic activity within the memory trace has significance for physiological explanations of behavior. If it is true that the trace tends to develop actively toward a more stable organization and not merely to decay with the passage of time, we are provided with important information about the nature of cerebral functioning. If, however, systematic errors of memory (which undoubtedly occur) are not due to an intrinsic activity of the trace but to the influence of external events upon it, we should rid ourselves of a conception which would mislead in the development of theory.

It must be recognized that the nature of the mnemonic trace is a central problem for any psychological theory that postulates real cerebral events as the basis of behavior. Two sets of facts are apparently opposed to each other. On the one hand, the persistence of early learning throughout life even despite major cerebral disturbances—including coma, anesthesia, *grand mal* convulsions and deteriorative diseases—argues strongly that some permanent and unmodifiable change in neural structures is the basis of the trace. On the other hand, the facts of perceptual generalization (**6**, **9**) and the apparent lack of specificity of most cortical tissues in learned behavior (**7**, **8**) seem to show that the trace cannot be a property of any particular set of cells; that is, that the trace cannot be structurally determined. On this rock psychological theories have foundered, or for fear of it have hardly dared leave port. The nature of the mnemonic

trace is a question of the first importance, and no source of information about it can be neglected. For this reason we attempt in the present paper to provide evidence by which the Gestalt hypothesis can be better evaluated.

As the reader will discover, our evidence leads to the conclusion that the hypothesis is wrong. We should like to make it clear, however, that this does not constitute an attack upon Gestalt theory as a whole. We reject the conception of autonomous changes of the trace only in urging that another dynamic mechanism, which is recognized as being in harmony with the broader aspects of Gestalt theory, is a sufficient explanation of the experimental facts, and more consistent with them.

## The Theoretical Alternatives

There has been confusion in the minds of both critics and supporters of the Gestalt hypothesis, concerning the theoretical issues involved and the significance of certain facts for the proof or disproof of the hypothesis. We shall first deal with these points.

According to Koffka (**5**, pp. 492, 498) the hypothesis of spontaneous activity within the trace originated with Wulf (who published in 1922). Wulf distinguished three classes of mnemonic error: 'normalizing,' 'emphasizing,' and 'autonomous' changes of memory. From Koffka's discussion it appears that these three spring from two mechanisms only. One mechanism is the operation of processes external to the trace—either changing the trace itself, or modifying the response which the trace otherwise would determine completely; this mechanism we shall refer to as that of 'external influence' on the trace. The other mechanism is the operation of stresses within the trace, the postulated tendency of the trace to reach a more stable organization; this is the mechanism of 'dynamic activity' or 'spontaneous change.'

We must realize therefore that Gestalt theory has from the first recognized, even emphasized, more than one source of mnemonic error. It differs from other theories in postulating a mechanism of spontaneous change within the trace, and denies that all errors can be reduced to the single mechanism of external influence, but it does recognize that external influences have an important effect. The question at issue consequently is not whether systematic errors of memory occur (that they do is recognized by both sides) nor whether all errors can be attributed to spontaneous change in the trace (this is denied by Gestalt theorists as much as by their opponents), but whether in explaining mnemonic errors we need to postulate two mechanisms, or only one.

Let us now look more closely at the way in which each of these two possible mechanisms would operate.

## Postulated Activity within the Trace

The Gestalt hypothesis of dynamic activity within the trace assumes that the trace is a more or less unstable organization of field forces which tends to shift toward equilibrium. The forces or tensions within the trace derive from similar forces which are supposed to operate in perception. Certain patterns of physiological activity, aroused in the brain by sensory excitations, are thought of as being in equilibrium; others as being in disequilibrium internally, and maintained only by the agency of the continuing sensory influx. When the sensory activity stops, the persisting process (or trace) at once begins a slow change toward internal equilibrium. Thus some visually perceived patterns are 'good' figures, arousing stable cerebral processes and laying down a stable trace. Imperfect figures lay down traces which tend to change toward equilibrium—that is, toward the characteristics of the trace that would be left by a 'good' figure. With the passage of time, consequently, the subject will remember a 'better' figure than the one he was actually shown: one that is nearer to being closed, more symmetrical, and so on.

But all this is supposed to apply only if the process is not disturbed by other events. The possibility of a strong influence upon the organization of the trace or upon its control of later responses was discovered by Wulf and by Gibson (**2**), and unequivocally demonstrated by Carmichael, Hogan and Walter (**1**) and Zangwill (**12**). Koffka explicitly states that the autonomous activity of the visual trace may be interfered with, particularly when there is an interaction between two visual traces or between a verbal and a visual trace. The Gestalt hypothesis does not require a change of memory, such that a better figure is reproduced or recognized, in every instance. It does, however, maintain that the tendency of the trace toward equilibrium is frequently dominant over external influence, and that there are errors of memory which can be explained in no other way. Koffka stresses particularly that internal forces must lead to a slow, long-continued mnemonic change, and treats such changes as the crucial evidence for the necessity of the Gestalt postulate of stress within the trace.

## External Influence from Another Trace

Koffka (**5**, pp. 501–503) makes it obvious that the actual response of the subject, in a memory test, must usually be influenced by more

than one trace. No one, however, has discussed the full implications of this fact as it is related to tests of the Gestalt hypothesis.

An interaction between two such traces might operate in a way that would produce a slow and directed increase of mnemonic error. Let us suppose that we have two traces, A and B, laid down at the moment of perception. One might be visual, the other verbal—the subject looking at a triangle, let us say, and at the same time thinking of the word 'triangle.' The visual trace, A, must be more determinate (as related to a particular triangle) and at the same time more subject to decay, than the verbal trace, B: one can remember the word 'triangle' longer than one can remember the exact visual proportions of a particular figure. Immediately after seeing the figure, the *S*'s recognition or reproduction of it would be dominated by the more determinate visual trace, A; but as A begins to fade a point would be reached at which B would assume complete control. When the verbal trace, B, is alone effective the *S* on being asked to reproduce the figure would presumably draw a 'stylized' triangle—remembering only that it was a triangle and not its particular characteristics (*cf.* **11**, p. 82). We do not know whether the shift of control from A to B would be gradual, both tending to affect all responses but to a varying degree, or whether it would be sudden and complete. In the one case, there would be a steady increase of error with the passage of time. In the other, error would be delayed and sudden.

With complex figures, as Koffka shows, there might be two visual and thus two verbal traces, and the complications of possible interaction between them are beyond the range of useful speculation. Almost any temporal pattern of mnemonic change could be produced. This applies also when there are a number of figures to be remembered simultaneously (since there is repeated indication of interaction in the experiments that have used such a series of figures). Evidently, then, we must agree with Koffka's recommendation that mnemonic errors be studied in conditions as simple as possible. Such conditions are not to be found in any previous investigation of the problem.

## Evaluation of the Existing Evidence

*Experiments reviewed by Woodworth.*—The earlier attempts to demonstrate the truth of the Gestalt hypothesis of spontaneous activity within the trace have been reviewed critically by Woodworth (**11**), who shows that the evidence is unsatisfactory as proof and argues that it tends instead toward disproof of the hypothesis. But the very considerations which invalidate the earlier proofs affect Woodworth's suggested disproof also.

The principal flaw in the earlier evidence is that it was obtained by a method of repeated reproduction. The *S* is shown a diagram and asked to reproduce it after varying intervals. Hanawalt (4) showed that the kind of error which appears in a delayed reproduction, and which has been cited in support of the Gestalt hypothesis, is found also in direct copies made while the diagram is still before the *S*. Hence the errors of delayed reproduction may have no mnemonic origin. Further, when the test is repeated, what the *S* may remember is his own erroneous drawing in the first test. Error might thus become cumulative whether the trace, when left to itself, is static or dynamically active. It follows that the evidence of repeated reproductions cannot be used to support the Gestalt hypothesis.

But neither can it be used to attack the hypothesis. If the successive drawings do not become clearly 'better' figures, their failure to do so may be due to the interfering influence of traces left by the *S*'s own successive drawings, and not to a lack of dynamic activity within the trace. This point will become clearer as we discuss the data reported by Seidenfeld, who used repeated tests of recognition instead of reproduction.

*Seidenfeld's experiment.*—Seidenfeld (10) improved considerably on preceding work by using a recognition method, but unfortunately he continued the use of repeated testing. The *S* was shown a number of diagrams, and at varying intervals was asked to select each of the diagrams he had been shown from a corresponding test series. Each of these series included the correct (originally shown) diagram, with others which represented increasing degrees of error such as the experimenter was looking for. The first test made in this way is legitimate as no later test can be; on later tests does the *S* remember what he was shown originally, the diagram he selected on the preceding test, or one of the other erroneous diagrams?

We cannot suppose that the original trace is unaffected by the restimulation involved in testing, and two kinds of distortion are possible. Repeated stimulation with the *wrong* diagrams may induce errors that would not occur otherwise; on the other hand, repeated sight of the *correct* diagram could so reinforce an unstable trace as to prevent or limit spontaneous mnemonic change. The experiments that have used the method of repeated testing, either of reproduction or of recognition, are really inconclusive. Their results cannot be interpreted with respect to the question they were meant to answer.

Otherwise we should have to regard Seidenfeld's results as well-nigh conclusive evidence against the Gestalt hypothesis. He obtained marked reversals of the direction of error with the passage of time, which does not accord with the idea that the mnemonic changes

are determined by a single configuration of internal stresses within the trace. To make his results consistent with Gestalt theory Seidenfeld postulates a reversible gradient of dynamic change. But this explanation is not satisfactory, for it seems that it must refer to events external to the trace; and if the errors are a function of external events they are not an index of spontaneous change within the trace. In the analogies which Seidenfeld cites from physiology and chemistry, the reversal of gradient occurs only when some change is introduced into the system, and Seidenfeld himself mentions an interaction of verbal and other processes with the trace, as a cause of the reversals. The most probable cause of the reversals is found in the repeated stimulation with correct and incorrect diagrams; but in any event the experiment does not provide interpretable evidence concerning the existence or nonexistence of dynamic activity within the trace.

*Goldmeier's experiment.*—Goldmeier (3) has reported an experiment which at first sight appears to have overcome the difficulties of earlier investigations and to establish the Gestalt hypothesis beyond doubt. However, his evidence does not stand closer examination.

The experiment ingeniously avoided the difficulty of the reproduction method pointed out by Hanawalt, by measuring only the divergence of the *S*'s drawing from the same *S*'s copy of the diagram. Goldmeier's work is also the first to avoid the method of repeated testing. On the other side of the ledger, the details of procedure are not always clear, and the results which are most stressed in the report are definitely misinterpreted by the author.

Apparently each of the 162 *S*s copied each of six diagrams twice, and yet it is stated that only 468 copies were obtained. It is also not clear whether all six diagrams were presented to the *S* simultaneously. If, as it seems, they were not, we do not know whether the test 'immediately' after copying the diagram was made after all six had been copied, which would mean a variation of from one to perhaps 20 min. in the time interval described as 'immediate'; or whether the immediate reproductions were actually immediate, in which case they would be made when the *S* had only a single diagram to remember (the one he had just copied) and would not be comparable with later reproductions when the *S* was obliged to recall a number of diagrams simultaneously.

As to results, we must at once reject the only apparently clear-cut data which Goldmeier emphasizes most in discussion. In the tests with diagrams 1a, 1b, 2b, 4 and 6, what Goldmeier reports is the *frequency* of error of any kind or degree, and treats this as an index of the degree of directed or systematic error, which is of course quite illogical. Fading of the trace alone would produce a steadily in-

creasing frequency of random error in group scores. All-or-none changes and reversals of trend would not be revealed by this method of analysis. The data from these tests consequently are wholly irrelevant to the question at issue. What they show is that mnemonic error is more likely with the passage of time, which does not need a special experimental demonstration.

The errors made in other tests (2a, 3a, 3b, 5a and 5b) are not a sufficient evidence of dynamic activity within the trace. Only in one test (3b) is there a constant increase of error in one direction. Since the differences are slight and not statistically significant, and since the apparent trend is found with only one out of five test diagrams, the trend could readily be only a chance distribution of error scores in the four tests made with this one diagram. With the other four diagrams errors were static or subject to slight reversals during a significant part of the total six-week test period.

Goldmeier's data therefore provide no more adequate support for the Gestalt hypothesis than those of preceding experiments.

## Conditions for Testing the Gestalt Hypothesis

Nowhere do we find the evidence that makes it necessary to postulate a special mechanism of spontaneous dynamic change within the trace. The belief that there is such a mechanism however is well established, and no one has as yet met the need referred to by Koffka (**5**, p. 503) for a test of the hypothesis in circumstances as simple as possible. It is evidently worthwhile to make a further experimental investigation in conditions which will favor the appearance of interpretable mnemonic errors in accord with the hypothesis.

What are these conditions? First, difficulties of interpretation require the avoidance of repeated testing. Secondly, the method of reproduction should be avoided also. In spite of Goldmeier's ingenious solution of the major difficulty of the use of reproductions, he found himself obliged to fall back on a rating method to deal with some drawings, and Koffka's discussion no less than Woodworth's points to recognition as the best index of the state of the trace. Thirdly, it is desirable to avoid the use of complex diagrams, of ones which are at all similar to each other, or of a number of diagrams to be remembered by the same *S*s. One must agree with Koffka that these are conditions which would make for an interaction between traces and tend to obscure any spontaneous change within the individual trace. Finally, in using the recognition method, one must choose test diagrams which will lead to a predictable mode of mnemonic change; and the change should be in a continuum in which the various steps cannot easily be defined in words, in order to minimize any interference of verbal with visual trace. This last point is elaborated below.

## Choice of Test Diagrams

In arranging experimental material, we first defined two series of diagrams for recognition tests, and then selected two from each as the diagrams to be shown to the *S*s for their later recognition. One series was made up of circles with gaps of increasing size; the other of arrow head designs with decreasing sharpness of angle.

(1) Twenty-four circles, with gaps which increased by steps of 2½° from a gap of 2½° to one of 60° (Fig. 1) were drawn with india ink on separate 7.5-cm. squares of white bristol board drawing paper. Each circle was 5.6 cm. in diameter. Nos. 8 and 16 from this series (20° and 40° gaps) were duplicated on identical pieces of bristol board. These were for use as the original stimuli, one of which each *S* would see and be asked to remember.

(2) Twenty-four arrowhead figures were prepared as a test series in the same way. The variable here was the sharpness of the point of the arrow, the distance between the barbs being constant, as well as the length of the horizontal line joining the point of the arrowhead to the centre of its notch. The mode of variation is best comprehended from Fig. 2, from which it can be seen that the series varied from a very pointed arrowhead with deep notch (no. 1) to a laterally unsymmetrical quadrilateral (no. 24).

The circle-with-gap series was qualitatively continuous throughout, the differences between any two successive items being definable in words only by making an estimate of the linear or angular extent of the gap. This fact is clearer if one envisages the effect of extending the series to include a circle with a gap of 180°: this would be qualitatively distinct from the smaller gaps, and verbally definable as a semicircle and not merely as a circle with a very large gap.

In the arrowhead series, there was a definite point of qualitative inflection with diagram no. 17, where the notch became a straight line; and at no. 21, which was a laterally and vertically symmetrical diamond. This series above no. 17, therefore, was not suited to reveal the existence of spontaneous changes of the trace uninfluenced by external processes. From nos. 1 to 16, however, it appears that there was qualitative continuity and that differences between items could be verbally described only by an estimate of proportions or of angular extent.

Our choice of these figures was determined by their use in previous investigations (though we have added a horizontal line in the arrowhead figure), their relative simplicity and their marked dissimilarity to each other, as well as by the fact that for each a mode of mnemonic variation could be predicted from Gestalt theory, in which qualitative homogeneity would not be inconsistent with fairly wide quantitative differences. The *S*, being asked to remember one of the diagrams, may say and remember 'circle with a small gap' or 'arrowhead, sharp,' but these verbal cues would not be decisive in the later selection of a single diagram and over a rather broad range of possible mnemonic change would not prevent the *S* from choosing the particular diagram which corresponded to the present state of his *visual* trace. A verbal trace might be decisive in the selection if an accurate estimate of linear or angular extent were made at the time of observing the original figure (the one to be remembered); but since nothing was said to the *S* about mode of variation or method of testing recognition, it is very unlikely that more than a few *S*s would make an accurate verbal description of the figure in quantitative terms. If we concede the possibility that some might do so, and that this might obscure the effect of the postulated dynamic activity within the visual trace, there would still be another group in which the verbal trace would not interfere with the mnemonic visual changes. The distribution of errors of recognition would then give a cue to the presence of disturbing verbal traces, by taking a bimodal form. Further, the bimodality would become more marked with the passage of time.

At best, this does not eliminate difficulty in interpreting the experimental results, but there seems to be no way of setting up circumstances more favorable to the demonstration of the Gestalt hypothesis; and Wulf and Koffka state clearly that in favorable circumstances the operation of internal stresses in the trace will override the disturbing effect of coexistent traces or other processes, including verbal ones.

Finally, use of two separate diagrams in each series as original stimuli, with separate groups of *S*s, achieves two objectives. If verbal cues affect the recognition of one of the two diagrams, they may not be as effective with the other. Also, if the particular figure chosen as original stimulus happened to be close to the 'perfect' instance of such a figure, Gestalt theory would not require that progressive error occur. By using two diagrams from each series the probability is increased of finding, for one figure at least, conditions suitable to the demonstration of spontaneous activity within the trace.

### Subjects

Two separate experiments with different *Ss* were made. In the first the *Ss* were 150 adult Summer School students, chiefly teachers, at Queen's University; in the second, 255 Junior High School students 11 to 16 years of age, 90 percent of whom were between 12 and 15.

In each experiment the *Ss* were divided into '5-min.' and '24-hour' groups, assigning individuals alternately to each group as their consent to act as *Ss* was obtained. Half of each group was shown diagram A, and either C or D; the other half diagram B and either C or D. The choice between C and D was made in such a way that approximately the same number of subjects were exposed to each of the following combinations: A, C; A, D; B, C; B, D.

### Procedure

Each *S* was worked with individually, first with one of the circle-with-gap diagrams (A or B) and then with one of the arrowhead diagrams (C or D). In the first experiment, with adult *Ss*, none of the tests overlapped in time. The stimulus was shown and recognition tested with a circle diagram, and the procedure was repeated with an arrowhead diagram. There seemed however to be no possibility of interference effects between diagrams so dissimilar, and in the second experiment with school children the 24-hour *Ss* were shown a circle diagram for five sec., and an arrowhead diagram for five sec., on the same day; and were tested on the following day. We can find no evidence that the change of procedure affected our results. Although there are differences of performance between adults and children, the differences are as evident in the 5-min. groups, where children and adults were tested in exactly the same way, as in the 24-hour groups with whom the difference of procedure occurred. As far as they affect the conclusions of the study, the results were the same with adults and children.

The *S* was first told that he would be shown a diagram which he was to remember exactly, and that the accuracy of his memory would be tested. The stimulus figure was then shown him for five sec. At the end of the 5-min. or 24-hour period he was handed the 24 recognition patterns (the test series described above), each on a separate piece of bristol board as the original stimulus had been, and told to pick out the one which he had been shown.

The pieces of cardboard bearing the test diagrams were strung on a metal ring, in order from no. 1 to no. 24. However, if the bundle thus held together were 'opened' for example at no. 10, nos. 1 to 9 would move around the ring and be behind no. 24—turning the '1–24' series into a '10–24–1–9' series. To avoid a constant error from having no. 1 always on top, the cards were so presented that some card from no. 1 up to the 'correct' pattern alternated, as top card, with some card from the correct pattern to no. 24, randomizing the choice of the particular card within each part of the series.

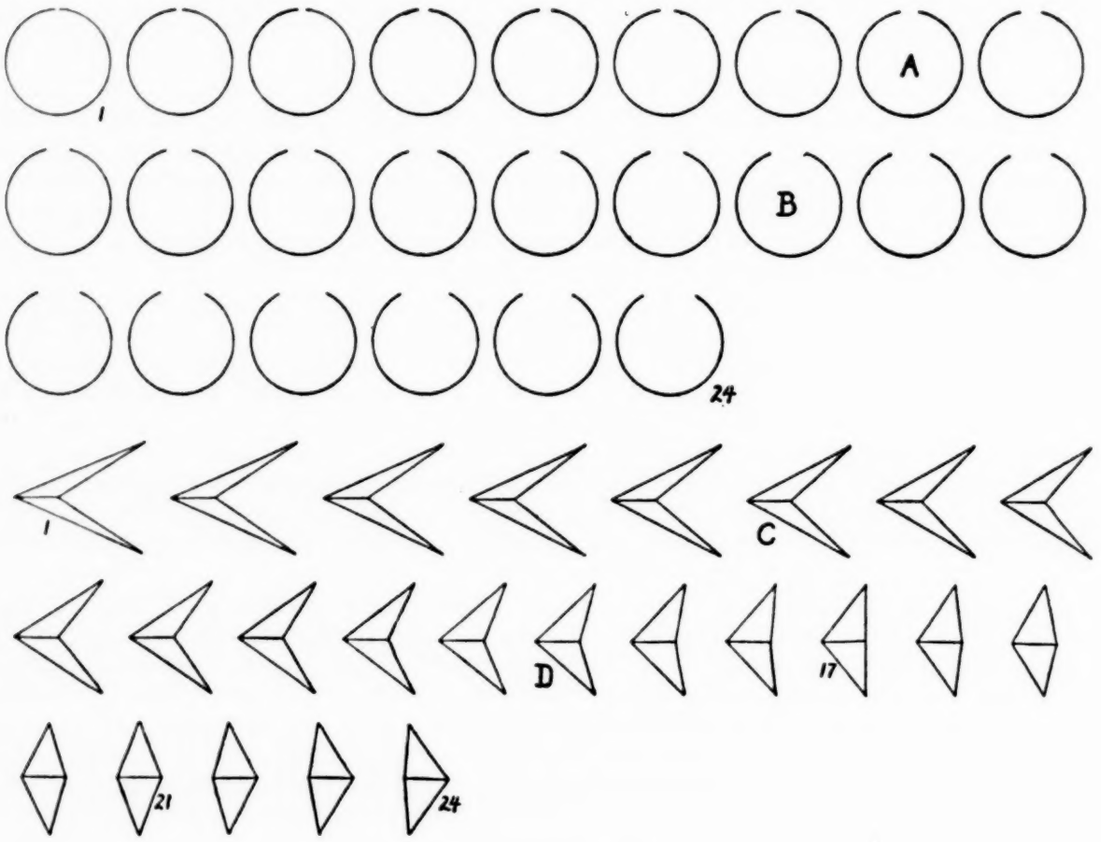
## Results

The distributions of errors made by adults and children are somewhat different, so that we cannot combine the data of the two experiments. The children's errors have a greater spread, and the mean error for adults and children is significantly different in each of the following tests: Circle A (small gap: see Fig. 1), 5-min. period; Circle A, 24-hour period; and arrowhead D (blunt: see Fig. 2), 5-min. period.

Table I presents the distributions of errors of recognition, in terms of minus or plus deviations from the original stimulus. Table II gives the mean error in each test, indicates the statistical reliability of the difference from zero, and shows which differences are 'progressive'—that is, in which experiments the mean error was greater, in the same direction, after 24 hours than after five min. The distributions are grouped in Table I so that their trends can be

distinguished, but statistical constants were calculated from the ungrouped scores.

In all, 16 tests of recognition were made, 8 with children and 8 with adults. Seven of the 16 mean error scores are statistically significant as deviations from a zero error in recognition. But there is no agreement between child and adult experiments as to which




FIGS. 1 and 2. Circle-with-gap and arrowhead series used to test recognition after showing the *S* a duplicate either of item A or of B, and a duplicate of C or of D. The items are numbered from left to right in each figure, A being number 8, for example, C number 6, in their respective series. The diameter of the circles of Fig. 1 was 56 mm. The vertical extent of the arrowhead diagrams of Fig. 2 was 60 mm.

stimulus figures lead to such error. For the adults, significant error occurs with the arrowhead figures, C and D; for the children, with the circle figures, A and B.

The 16 tests permit 8 comparisons of the amount and direction of error in 24 hours as against five min. Of these 8, 5 are 'non-progressive.' Only 3 are 'progressive.' But none of the differences between 5-min. error and 24-hour error in any of these comparisons is statistically reliable, and the *t*-values range from the very small up to one that approaches but does not reach the five percent level of confidence—precisely as would be expected if the 5-min.: 24-hour differences are a product of chance errors of measurement. We have

TABLE I

ERROR DISTRIBUTIONS IN THE RECOGNITION OF FOUR PATTERNS, A TO D, SEPARATELY FOR CHILDREN AND ADULTS AND FOR 5-MIN. ('5') AND 24-HR. ('24') PERIODS

The Number of Correct Recognitions (within One Step of the True Pattern) is Italicized in Each Distribution. A Minus Sign Means an Error in the Direction of Pattern No. I. *N*, Number of Subjects in Each Subgroup.

| Amount of Error | Circle A (No. 8) | | | | Circle B (No. 16) | | | | Arrow C (No. 6) | | | | Arrow D (No. 14) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Child | | Adult | | Child | | Adult | | Child | | Adult | | Child | | Adult | |
| | 5 | 24 | 5 | 24 | 5 | 24 | 5 | 24 | 5 | 24 | 5 | 24 | 5 | 24 | 5 | 24 |
| −13 to −11 | | | | | | | 1 | | | | | | | 1 | | |
| −10 to − 8 | | | | | 2 | 2 | 2 | 1 | | | | | | 1 | | |
| − 7 to − 5 | | | | | 2 | 3 | 2 | 5 | 14* | 7 | 1 | 2 | | 3 | | 2 |
| − 4 to − 2 | 4 | 4 | 6 | 8 | 8 | 13 | 5 | 5 | 12 | 12 | 13 | 13 | 9 | 7 | 12 | 9 |
| − 1 to 1 | *15* | *20* | *20* | *14* | *24* | *15* | *18* | *11* | *14* | *23* | *20* | *12* | *49* | *45* | *22* | *23* |
| 2 to 4 | 21 | 20 | 10 | 7 | 16 | 13 | 5 | 14 | 20 | 13 | 5 | 8 | 4 | 4 | 2 | 1 |
| 5 to 7 | 8 | 12 | 1 | 3 | 10 | 9 | 3 | 3 | 1 | 8 | | 2 | 1 | 0 | | 1 |
| 8 to 10 | 8 | 4 | | 2 | 2 | 6 | 1 | 1 | 1 | 1 | | | 2 | 3 | | |
| 11 to 13 | 5 | 4 | | 0 | | | | | | | | | | | | |
| 14 to 16 | 3 | 2 | | 1 | | | | | | | | | | | | |
| *N* | 64 | 66 | 37 | 35 | 64 | 61 | 37 | 40 | 62 | 64 | 39 | 37 | 65 | 64 | 36 | 36 |

* An end effect: − 5 was the largest error possible, occurring with the choice of diagram no. 1.

TABLE II

DIRECTION AND AMOUNT OF THE DEVIATIONS OF MEAN RECOGNITION SCORES FORM THE TRUE VALUE SEPARATELY FOR CHILDREN AND ADULTS AND FOR 5-MIN. AND 24-HR. PERIODS

'P' Indicates a Progressive or Consistently Increased Error in 24 Hours as Compared with 5 Minutes, 'NP' a Nonprogressive Change; '5%' or '1%' Indicates That the Deviation of the Mean from *Zero* is Significant at That Level of Confidence. The Unit of Measurement is the Step or Interval between Any Two Successive Diagrams of the Test Series.

| Stimulus | Interval | Children | | Adults | |
|---|---|---|---|---|---|
| Circle A | 5 min. | +4.33(1%) | NP | +0.30 | P |
| | 24 hr. | +3.58(1%) | | +1.26 | |
| Circle B | 5 min. | +1.03(1%) | P | −0.32 | NP |
| | 24 hr. | +1.13(5%) | | +0.60 | |
| Arrow C | 5 min. | −0.31 | NP | −0.97(1%) | NP |
| | 24 hr. | +0.50 | | −0.27 | |
| Arrow D | 5 min. | +0.14 | NP | −0.69(5%) | P |
| | 24 hr. | −0.31 | | −0.88(5%) | |

tabulated the 'progressive' and 'non-progressive' changes as such in order to show that they also occur as if by chance and are consistent with the idea that there is no real difference in 5-min. and 24-hour scores. Nothing about our results suggests that larger groups of *Ss*

would have given statistical reliability of the 5-min. : 24-hour differences. The data must therefore be interpreted as showing that the mean error score is the same after 5-min. and 24-hour intervals.

It should be noted that our results show no increased bimodality with the passage of time. This is evidence both against conflict of external influence with internal dynamic activity of the trace, and also against individual differences in the direction of internal dynamic activity. As shown above, if an external verbal trace were obscuring the action of internal stresses in the visual trace, this fact should be revealed, with the particular material of the present study, in bimodality of distribution which would be more marked with greater time intervals. Bimodality is suggested in one distribution only, with Diagram C, children, 5-min. subgroup. This is presumably an end-effect and would not have appeared if the test series had been extended in the direction 3–2–1–. . . . At any rate, the bimodality cannot be treated as supporting the idea that a conflict of internal and external forces occurs with this diagram, for there is no bimodality in the distribution of children's errors after 24 hours, and not the slightest bimodality of adult scores with the same diagram.

Since Gestalt theory holds that spontaneous changes within the trace may move toward various conditions of stability, an average error may fail to reveal systematic changes which in different individuals are in different directions. In our series changes of the trace might be toward either opening or closing a gap, sharpening or levelling an angle. But these opposed trends would lead to a bimodal distribution of scores, more marked with the passage of time, and the absence of such a pattern is evidence against the existence of the opposed dynamic trends.

## Conclusions

In brief, our data show the occurrence of systematic errors of memory which are at a maximum in five min. or less and then become static for an undetermined period greater than 24 hours. The lack of evidence of a progressive increase of error is apparently not because we had too small a number of *S*s. There is every indication that larger groups of *S*s would not have given a different result. It must be concluded that memory for such figures as we have used does not show the steady, directed change that is required by the Gestalt hypothesis.

It has been seen that the true question in evaluating the Gestalt hypothesis is whether we need to postulate two sources of mnemonic error, or only one. Are there errors which cannot be explained by the agency of what we have called external influence, and which

require the extra assumption of spontaneous activity within the trace itself? It has also been seen that there is no evidence from the preceding experiments by other workers which makes the assumption necessary. The earlier evidence obtained by repeated reproductions is not valid, and the data of Seidenfeld (**10**) and Goldmeier (**3**) offer no clearer support for the Gestalt hypothesis. It appeared to us, in planning the present experiment, that an investigation should be made in circumstances which *a priori* would favor the appearance of the postulated dynamic activities of the trace. But our data also fail to support the hypothesis, and it must be concluded that there is no evidence whatever that makes it necessary to postulate two distinct mechanisms of mnemonic error, as the Gestalt hypothesis does.

But we believe that we should go father in our conclusions and argue that our experiment definitely controverts the hypothesis. To make our results consistent with it one must assume that some external influence on the trace has interfered with the postulated internal activity. In discussing our choice of diagrams for study, we have shown the probability that such a conflict between forces internal and external to the trace would, in the conditions of our experiment, be revealed in an increasing bimodality of the error distribution. Such a bimodality is not found in any of the tests, with four different test diagrams. Also, Gestalt theorists have stated definitely that in favorable circumstances the postulated spontaneous activity of the trace will dominate external influence (otherwise the spontaneous activity could never be discovered or demonstrated). We do not see what circumstances could be much more favorable than those of our circle-with-gap tests. The actual fact is that since the obviously unsatisfactory method of repeated reproduction was abandoned no experimenter has observed the slow, persistent and directed increase of mnemonic error which Koffka treats as the crucial evidence for the Gestalt hypothesis. Not only our results, but also those of Goldmeier as far as they are relevant to the question, seem to be a positive refutation of the Gestalt hypothesis.

## Summary

Two experiments, using a total of 405 *Ss*, were carried out as a test of the Gestalt hypothesis of dynamic activities within the trace. The methods of reproduction and of repeated testing of the same *Ss* were avoided in favor of recognition tests with different groups for different time intervals. Results failed to support the theory, and a review of the literature shows that the progressive, directed increase of error required by it has been found only in experiments using the

invalid test method of repeated reproductions. It is concluded that there is a complete lack of factual support for the Gestalt hypothesis of spontaneous activity within the trace.

(Manuscript received March 5, 1945)

Bibliography

1. Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.
2. Gibson, J. J. The reproduction of visually perceived forms. *J. exp. Psychol.*, 1929, **12**, 1–39.
3. Goldmeier, E. Progressive changes in memory traces. *Amer. J. Psychol.*, 1941, **54**, 490–503.
4. Hanawalt, N. G. Memory trace for figures in recall and recognition. *Arch. Psychol.*, 1937, No. 216. Pp. 89.
5. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.
6. Köhler, W. *Gestalt psychology.* New York: Liveright, 1929.
7. Lashley, K. S. *Brain mechanisms and intelligence: a quantitative study of injuries to the brain.* Chicago: Univ. Chicago Press, 1929.
8. Lashley, K. S. Coalescence of neurology and psychology. *Proc. Amer. phil. Soc.*, 1941, **84**, 461–470.
9. Lashley, K. S. The problem of cerebral organization in vision. In: *Biological Symposia*, 1942, **7**, 301–322 (edited by H. Klüver).
10. Seidenfeld, M. A. Time as a factor in the recognition of visually perceived figures. *Amer. J. Psychol.*, 1938, **51**, 64–82.
11. Woodworth, R. S. *Experimental psychology.* New York: Henry Holt and Co., 1938.
12. Zangwill, O. L. A study of the significance of attitude in recognition. *Brit. J. Psychol.*, 1938, **28**, 12–17.

# REASONING IN HUMANS

## III. THE MECHANISMS OF EQUIVALENT STIMULI AND OF REASONING[1]

BY NORMAN R. F. MAIER

*University of Michigan*

### Introduction

The concept of equivalent stimuli, developed by Klüver (**3**), has contributed much to our understanding of the problems of discrimination and perception, but the implications of this concept may be even more significant in the field of the thought processes, as is suggested by the work of Halstead (**2**) and Strauss and Werner (**14**). Maier and Schneirla (**13**) have pointed out that the mechanisms making for equivalent stimuli can adequately account for transfer of training and abstraction behavior observed in experiments with animals below man. If a second situation contains the features to which the animal has learned to respond in the first situation, then the second situation should call out the learned response.

Suppose an animal is trained to choose a white card having a black circle 6 inches in diameter whenever this card accompanies a similar card with a 3-inch circle. If it reacts to the aspect 'larger' then this previous experience will cause it to choose a 12-inch circle when it accompanies the 6-inch circle, but if it reacts to the circle of a given size then it will choose the 6-inch circle no matter what accompanies it. However, the animal also may choose the darker of the two cards in the training pair. In this case the card with the larger black circle would be selected because it is the darker, but the choice would correspond to that made on the basis of size. Brightness and size reactions could be distinguished if white 6- and 3-inch circles were put on black cards. In this situation a size reaction would call for the choice of the card with a 6-inch circle and a brightness reaction would call for the card with the 3-inch circle. Tests of this type show how the nature of the animal's learning in the first situation will determine its reactions to new pairs of cards. If we want the animal to choose on the basis of size or brightness or any other feature, then we must make sure that the training is such as to produce this learning. Merely training the animal in the first situation and obtaining con-

[1] Mr. Harold Guetzkow generously assisted the author in this experiment.

sistent (correct) reactions does not necessarily accomplish this end (11).

It is possible that this formulation of transfer in training is adequate for explaining all problem solving which is more complex than random trial and error. In such a case reasoning becomes an empty term insofar as it is used to designate a kind of problem solving mechanism. Whether or not a person solved a given problem would depend upon *what* he has learned previously, rather than upon his ability to apply his learning to new situations. The same training would permit different people to learn quite different things, so that success in subsequent problem situations would vary from individual to individual because of this difference in what is learned, rather than because of differences in ability to reason. In order to improve success in problem solving it would then be necessary to place emphasis on training so that individuals will learn those things which will make the proper future situations equivalent to the training situation.

On the other hand, when reasoning is defined as the ability to combine isolated experiences in such a manner as to overcome a difficulty (Maier, **4**, **5**, **6**, **7**) we imply that past experiences can be reorganized and changed. If this sort of thing occurs, it means that the stress of a problem situation can alter previous experiences, which then cease to be fixed in their function. From this point of view success in reasoning will not be limited by the way we have learned things, but will depend upon the readiness with which the past learning is subject to modification and reorganization.

Thus, if a problem demands that a pair of pliers be used as a pendulum bob, the equivalence notion applied to problem solving demands that in our experience with pliers we have reacted to its weight aspect as well as to its other properties. The author's definition of reasoning, however, implies that even if an individual has never reacted to the weight aspect, this property of the pliers might emerge during the stress of the problem situation. In such a case the change produced by reasoning might create a new set of equivalent stimuli and many objects having weight become equivalent. However, the equivalent stimuli in this instance are the product of reasoning and not the mechanism which solved the problem.

There is no doubt that equivalence reactions are adequate for solving some problems, and solutions of this sort may be regarded as the product of *reproductive* thinking. The question we wish to raise is whether this is the whole of problem solving, or whether, in addition, there are *productive* solutions produced by the reasoning process. This question is not too easily decided by direct experimental procedure and must be approached somewhat indirectly. We have assumed that problem solving by the mechanism of equivalent stimuli

implies that what is learned does not change after the learning period has terminated, but what we learn can change as long as the learning period continues. In other words, memories of things remain relatively constant, whereas perceptions of the same things may change in that a second or third look may cause us to react to a new aspect. If this assumption is valid it follows that if a problem requires a modification in what is learned, then perceptions (which give an opportunity for relearning) should aid solutions but memories of these perceptions should not. Further, incorrect learning should be no more useful than no learning at all.

## Precedure

To test the relative value of perceptual and memory experiences in solving a problem, we divided our *S*s into three groups, each group containing 13 male and 12 female students from the introductory course in psychology. The *S*s worked individually and all were urged to keep the experiment a secret. The procedure for the experiment was as follows.

Group I. Each *S* was told that we wished to hang a couple of strings from the ceiling but that we could not deface the ceiling and therefore a little construction was required. With *E*'s aid the structures shown in Fig. 1 were then built. It will be seen that the pole on which

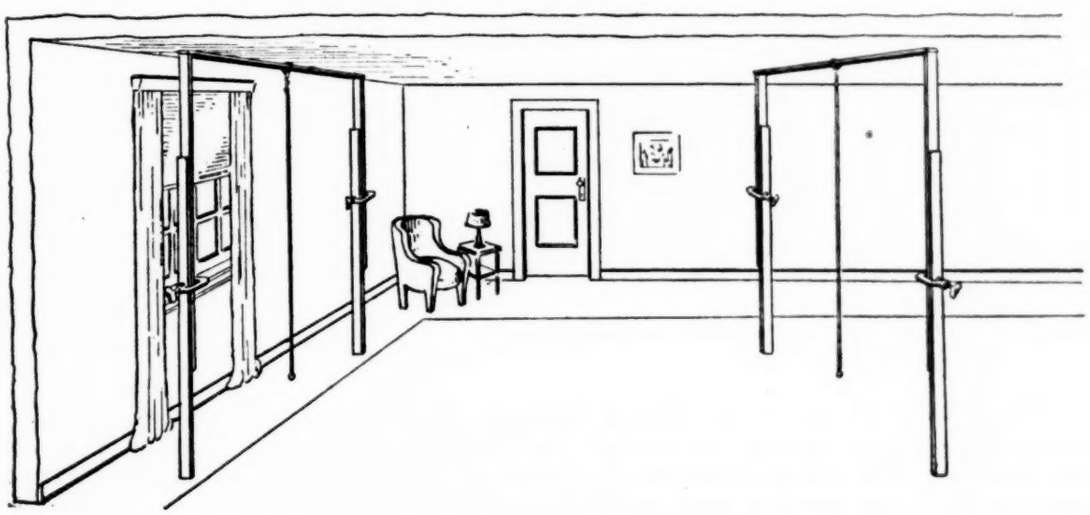


Fig. 1. The apparatus used in the String Problem. A string is suspended from each of the horizontal poles placed against the ceiling. Two upright supports hold each of the horizontal poles in place. Exact fit for the supports is obtained by clamping two poles together. The problem demands that the free ends of the strings be tied together. To do this one of the strings must be transformed into a pendulum.

each string was fastened was held in place with supports made by clamping together two shorter poles. When these structures were completed, the *S* was asked to tie the strings together. Since the distance between the strings was too great for a person to reach one while holding the other, the *S* was confronted with the problem of how to get hold of the ends of both strings in order to tie them together. This is the problem previously used by Maier (9, 10). The desired solution is that of making a pendulum out of one string by adding a weight. When the weighted string is put in motion the *S* is free to obtain the second string, carry the end toward the pendulum string and catch it as it swings toward him.

Since this problem was only to serve as a previous experience, the *S* was given enough time to appreciate the problem and was soon offered suggestions so that the solution was obtained within a 10-min. period.

When the solution to the string problem was obtained, the *S* was told that the purpose of the first problem was just to get him adjusted to the situation and that we would now give him another problem which he would have to solve by himself. The structures for the string problem were left standing and therefore, were always in plain view.

The *S* was then taken to another end of the room and asked to construct a hatrack in a certain spot. The available equipment consisted of two 1 inch × 2 inch poles (one 6 and the other 7 ft. long) and a 3-inch 'C' clamp. A hatrack was then defined as something sturdy enough to properly hold a heavy coat and hat. The problem was previously found to be very difficult (**10**), and is an adaptation of the problem used in my first investigation of reasoning (**8**).

The solution requires merely the clamping of the poles together and wedging them between the floor and ceiling. The solution is shown in Fig. 2.

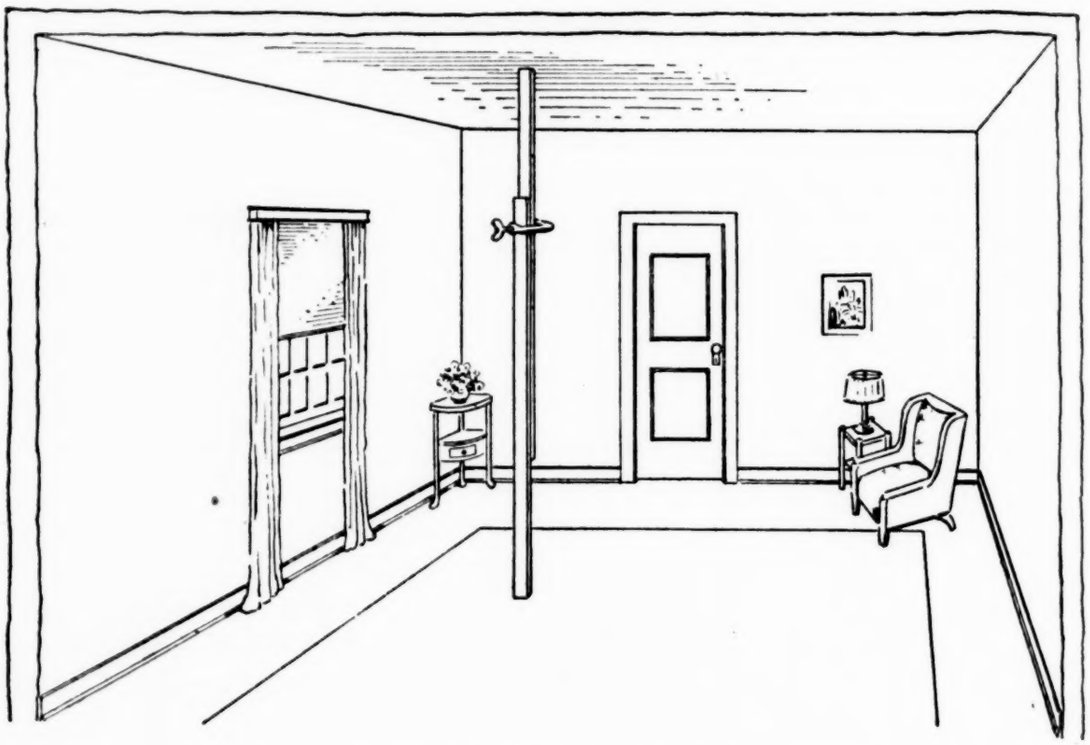


FIG. 2. The solution to the Hatrack Problem. The upright support is held in place by wedging it between the floor and the ceiling. The exact length is achieved by clamping two short poles together. The clamp serves also as a hook for hat or coat. Note that four hatracks appear in Fig. 1, but here they serve a different function.

It will readily be seen that the complete solution has already been experienced in constructing the apparatus for the string problem, but in this case the poles support something against the ceiling, whereas in the hatrack problem the ceiling supports the poles. Does one learn that the ceiling supports a pole when one experiences a pole functioning as a support? If so, there is a basis for equivalence; if not, then the equivalence of the situations can be questioned.

Group II. The procedure for this group was the same as for Group I, except that the string structure was taken down and the material removed from the room. The experience of the string structure is only a memory for this group of *S*s during their work on the hatrack problem.

Group III. This group served as a control for the other two groups. The *S*s were given no experience with the string problem and worked only on the hatrack problem.

All groups were given 30 min. to work on the hatrack problem. This is a generous amount of time for this problem, and *S*s who have not solved it in the allotted time are most willing to 'give up.'

As a further control for these three experimental groups, conditions were reversed for another group of 180 *S*s (55 men and 145 women). It was the problem of this group to discover what problem the hatrack solution solved. This group, therefore, was exposed to the experience

of seeing *E* clamp two poles together and wedge them between the floor and ceiling. As he worked he told them that he was constructing the solution to a problem, and it was their job to figure out what useful functions the structure could possibly serve. Each person was requested to write down all of the things or functions or uses he could figure out for the apparatus. No time limit was set.

If the solution of the string problem contains an element which is equivalent to the solution to the hatrack problem, then a presentation of the element itself (poles clamped together and wedged between floor and ceiling) ought to be experienced (among other things) as having a function similar to that of a hatrack. Certainly the string problem tends to obscure the function that the ceiling supports the pole, so that by presenting the element by itself we are, if anything, influencing the results to favor an interpretation in terms of equivalent stimuli.

## Results

Since a solution by equivalence demands that the function achieved in a solution be similar to a function previously learned, we will first discuss the results of the last group, which was requested to recognize the hatrack solution. The types of responses obtained are tabulated in Table I. By far the most common reaction is that of

TABLE I

Types of Functions Attributed to Hatrack Structure

| Function Attributed to Hatrack Structure | Percent Listing Function | | |
|---|---|---|---|
| | Men | Woman | Total |
| 1. Prop, brace or support, etc. | 80.0 | 74.5 | 75.6 |
| 2. Measuring device, to find distance between floor and ceiling, etc. | 14.3 | 29.0 | 26.1 |
| 3. Post, clothes-line pole, something to lean against, etc. | 17.1 | 25.5 | 23.9 |
| 4. Obstruction, division for room, etc. | 25.7 | 17.2 | 18.9 |
| 5. Adjustable boards, boards clamped for glueing, etc. | 11.4 | 8.3 | 8.9 |
| 6. Wedge | 5.7 | 9.7 | 8.9 |
| 7. Responses which were a carry-over from class discussions, spinal cord, synapse, etc. | 8.6 | 6.9 | 7.2 |
| 8. Extension | 5.7 | 4.1 | 4.4 |
| 9. Something to make people think | 18.8 | 1.3 | 4.2 |
| 10. For stretching curtains, pressing pants, etc. | 0 | 3.8 | 3.1 |
| 11. To hold stick in place | 0 | 2.7 | 2.2 |
| 12. Hang things on | 8.6 | 0 | 1.7 |
| 13. Hat or coat hanger | 0 | 8.3 | 6.6 |
| 14. Total of last three items | 8.6 | 11.0 | 10.5 |

recognizing the structure as one which supports something overhead. Thus, pillars support buildings or ceilings, but buildings and ceilings do not hold up pillars. This function of the poles as a support was found by 75.6 percent of the group. All other classifications are limited to less than 27 percent of the *Ss*. The hatrack function is distinctly apparent in the reaction of 10.5 percent of the *Ss*. In this classification we have included all responses which indicate that the stick is being held up by the ceiling or that the stick can serve as a

support from which things can be hung (items 11, 12 and 13). It may be argued, therefore, that, since 10.5 percent of the college population see the solution of the hatrack problem as a functional hatrack, it follows that no more than this percentage would be able to build the hatrack from equivalent past experiences. Even then we assume that they would have had the pertinent past experiences.

In order to defend the notion that all problem solving is a matter of equivalence some observers may believe that even the *post* interpretation (item 3) is equivalent to the hatrack, since the post involves the relation of a stick standing (by the use of ceiling) as does the hatrack. If this view is acceptable then 23.9 percent would solve by the process of equivalent stimuli. (Because all Ss who mentioned items 11, 12, or 13 also mentioned item 3 the value of 23.9 percent cannot be added to the 10.5 percent who mentioned items 11, 12 or 13.) The most generous interpretation of the data therefore would permit one to say that in 23.9 percent of the Ss the hatrack problem can be solved by a transfer of learning from the string problem. If a greater percentage than this solve the problem, then it must be conceded that something additional must be added to our explanation of problem solving behavior.

With this background of results on the recognition of the hatrack function when the solution is shown to Ss, we may turn to the results dealing with the phase of the experiment which demands the construction of the solution when the problem is presented. The results for the three groups are given in Table II.

TABLE II

SOLUTIONS TO THE HATRACK PROBLEM

| | | No. of Subjects | No. who Solved Problem | Immediate Solutions | Delayed Solutions | Failures |
|---|---|---|---|---|---|---|
| Group I. | Men..... | 13 | 13 | 8 | 5 | 0 |
| | Women... | 12 | 5 | 2 | 3 | 7 |
| | Total..... | 25 | 18 | 10 | 8 | 7 |
| Group II. | Men..... | 13 | 8 | 5 | 3 | 5 |
| | Women... | 12 | 4 | 2 | 2 | 8 |
| | Total..... | 25 | 12 | 7 | 5 | 13 |
| Group III. | Men..... | 13 | 5 | 0 | 5 | 8 |
| | Women... | 12 | 1 | 0 | 1 | 11 |
| | Total..... | 25 | 6 | 0 | 6 | 19 |

As previously found by the author (10), this type of problem brings out a marked sex difference. In each group the greatest number of solutions were obtained by men. Out of a total of 39 Ss who failed to solve the problem, 26 (66.7 percent) were women and 13

(33.3 percent) were men. This is in contrast with the results presented in Table I, in which the sex difference did not appear. As a matter of fact, such solutions as given in items 3 and 14 were presented slightly more frequently among women than among men. (Although the data in Table I were obtained after the war increased the proportion of women in psychology classes, it is improbable that this selective factor altered the relative problem solving ability of men and women student populations. Rather, it seems that the past experiences of men and women are sufficiently alike to permit similar equivalence reactions, and only when added factors are involved does sex become a pertinent factor.) Since the three groups presented with the problem of the construction of a hat rack were equated for men and women, the sex factor need not concern us further.

Considering the groups as a whole, we find that in Group I, for which the string structure was left standing while the *S*s were engaged with building a hatrack, 18 out of 25 *S*s solved the hatrack problem; in Group II, for which the string structure was removed after the demonstration, 12 out of 25 *S*s (48 percent) solved the problem; and for Group III, for which no preceding experience was introduced, only 6 out of 25 *S*s (24 percent) solved the problem.

It is apparent from these findings that (1) the experience of the string structure aids the solution of the hatrack problem, and (2) the immediate presence of the structure (perception) is more effective than a memory of the structure. Do these facts favor the theory that solutions are obtained by a transfer in experience or equivalent stimuli? In the control experiment it was found that the hatrack solution was equivalent to a hatrack in only 10.5 to 23.9 percent of the *S*s, yet if the theory of equivalence is correct then the results of Group I show that 72 percent of the *S*s find an element in the string problem which is equivalent to a hatrack. This discrepancy is too great to attribute to chance, and it is more reasonable to suppose that the requirement of building a hatrack changed the *prop* aspect of the string problem into a hatrack. Thus, the stress of the problem situation appears to alter perception and thereby achieve equivalence. In this case the equivalence experienced in excess of 23.9 percent must be regarded as due to a change in the experience of relationships, rather than to a mere transfer of experience from one situation to another.

If this interpretation is correct it follows that the removal of the string structure (which is capable of being equivalent to a hatrack) will make its parts less subject to modification. For the 'prop' element in the string problem to be of aid in more than 23.9 percent of the *S*s, the memory of the structure must change. Since 48 percent of the *S*s solved the problem under these conditions, there is evidence

in support of the thesis that a problem situation may alter memories, though not as readily as perceptions.

That 25 percent of the *S*s solved the problem without the aid of the string structure (Group III) demonstrates either that other past experiences were available for modification or that other experiences which were equivalent to the hatrack solution were already in the repertoire of these subjects. The former interpretation is consistent with the view that the problem situation can alter past experience, the latter argues that problem solving is purely a matter of transfer. To contend that 25 percent of the *S*s can recognize the constructed hatrack in their past experience because either 10.5 percent or 23.9 percent (depending on which value one chooses from the control experiment) recognized the hatrack in the control experiment would make it necessary for one also to account for difficulties encountered in explaining the results of Groups I and II.

It seems more appropriate to say that 25 percent of the *S*s can modify their uncontrolled past experiences so that a hatrack is constructed; that an additional 23 percent (total, 48 percent) can modify their past experience if a selected memory is added to their fund of experience; and that an additional 47 percent (total, 72 percent) can modify their past experience if an added perception is introduced into the background of experiences.

Further support of this interpretation is found in analyzing the time at which the solutions occurred. If the experience of the string structure contains an aspect which is equivalent with one needed in the hatrack problem, one should expect the solution to appear very early. If, however, the aspect must be developed under the stress of the problem, more time, on the average, should be required. We therefore have divided the solutions into *immediate* and *delayed* solutions. All solutions occurring within five min. were regarded as immediate, and a total of 17 such solutions were obtained. In 12 instances the correct solution was the first one offered and in the other five the correct solution was the second. In the case of the 19 delayed solutions which were obtained, several other solutions were first offered, and in no case did the correct solutions occur in less than 10 min.

It will be seen that all 6 of the solutions in Group III were delayed, and that 8 and 5 solutions, respectively, of Groups I and II were delayed. That the number of delayed solutions is roughly equal in the three groups may suggest that immediate solutions were produced entirely by equivalence reactions, and that delayed solutions required reorganization. However, on a relative basis this argument breaks down. In Group I, 8 out of 15 *S*s (53.3 percent) who did not exhibit immediate solutions showed delayed solutions; in Group II,

5 out of 19 *Ss* (27.8 percent) showed delayed solutions; and in Group III, 6 out of 25 *Ss* (24.0 percent) showed delayed solutions. Thus, Group I had the greatest percentage of the delayed solutions, and this group had the structure perceptually present; Groups II and III had about an equal percentage of delayed solutions, and they were alike in that they had to work from memories, either gained in the experiment or outside of it. If delayed solutions indicate that experiences have been modified, then the group that had the greatest opportunity for modifying its experiences (because pertinent elements were perceptually present) shows the change to the greatest degree.

Thus, if we grant that experience must be changed to solve some problems, the data for (1) the total number of solutions, (2) the immediate solutions, and (3) the delayed solutions are consistent. If we assume that equivalence reactions produce the solutions, we should demand a relatively smaller number of delayed solutions in Group I than in Group II. The data thus seem to favor the view that problem solving may involve a reorganization of past experience.

The equivalence concept, however, may still be considered a desirable explanation, since it would reduce problem solving to an aspect of learning. To bring further evidence to bear on the question additional data were obtained.

After some *Ss* in Group II failed to solve the problem they were asked to write down what they had done. This included a drawing of the string problem showing the structure in place. Of the nine *Ss* asked to do this, seven solved the hatrack problem when asked if they wanted a last try before leaving. All seven solved it almost immediately. Having the structure present in a drawing apparently permitted its meanings to be modified to meet the demands of the problem. *Ss* in Group I who failed to solve the problem were asked to walk about the room (and thus had to encounter the string structure). This did not help in any case, but when *E* hung his coat on the clamp, they showed characteristic insight behavior and were disgusted with failure to see the solution earlier.

## Discussion of Results

The concept of equivalent stimuli has added much to our understanding of the way in which past experience influences subsequent behavior. When old experiences influence the behavior in novel situations, it is because the new situation contains (for the individual in question) an element or relationship to which the person has previously learned to react. Thus, transfer of training is a matter of identical elements, but only in the sense that the elements are subjectively identical. Objective identity is wholly inadequate, as was shown in the present experiment. Before objectively identical ele-

ments could aid in the solving of the problem investigated in this study, it was necessary for the *S* to change his perceptions or memories so as to make for subjective identity. This change in past experience is a necessary preliminary to solving many problems, and it is only in the solving of such problems that one may say that reasoning has taken place. Confusion on this point has arisen because all problem solving activity has been identified with reasoning (productive thinking) and the product achieved (a solution) is used to characterize the process.

In order to understand the process of productive thinking, it is first necessary to distinguish between situations in which subjective identity in past learning and the demands of the problem are present at the outset, and situations in which the identity must be achieved by alterations in past experience. This distinction is implied when one uses the term *reproductive thinking* to designate problems solved by the process of equivalent stimuli and the term *productive thinking* to designate problems solved by processes which involve changes or a restructuring in past experience. Wertheimer (**17**) and Duncker (**1**) have emphasized the changes in meaning which accompany productive thinking, and Maier (**4, 5, 8, 9**) has presented evidence to show that modifications in functional meanings arise through new combinations of parts of past experience. Nevertheless, this distinction between productive and reproductive thinking has failed to influence the majority of psychologists when they deal with the subject of reasoning.

Another source of difficulty in the analysis of productive thinking arises from the fact that reasoning and concept formation are treated as if they were necessarily closely related. The product of reasoning may be a solution to a problem, but the solution to a problem is not necessarily the product of reasoning. Likewise the product of concept formation is a concept, which may be the solution to a problem, but a concept (if it is viewed as a common element) may be an expression of the mechanism of equivalent stimuli or the product of reasoning. Thus, concept formation may involve either the reasoning processes or equivalent stimuli. If the latter process is sufficient, concept formation is the product of discrimination learning. Experiments which deal with this form of concept formation often are purely investigations of equivalent stimuli. If other concepts are achieved by the process of reasoning (productive thinking), then such concepts should be differentiated from the former. Here again, failure to distinguish between process and accomplishment in concept formation has led to confusion. The importance of this distinction was clearly brought out in experiments by Maier (**12**) and Wapner (**15**), who found that brain injury increases the number of equivalent stimuli

but does not aid a higher level of concept formation. Werner (16) likewise stresses the importance of a distinction between process and achievement. A psychological process is a function, and similar functions may achieve different results while different functions may achieve similar results.

Finally, we must distinguish between equivalent stimuli produced by past experience and equivalent stimuli produced by reasoning. On the one hand, two situations may be experienced as equivalent because a second situation contains for the individual features to which he previously has learned to react. On the other hand, a second problem situation may not be equivalent to another until after the problem has been solved. Thus, after constructing the hat-rack one may find the structure equivalent to previous situations. In this case the experience of equivalence is a by-product of problem solving, not the process by which the problem was solved. When one gets a new 'direction' in reasoning, a new set of equivalent stimuli are created; and it is for this reason that 'direction' in thinking serves a selective function in thinking (8, 9).

When Klüver (3) found that a monkey would use a sack, a piece of rope, a piece of wire netting, a chain, and a cardboard triangle to pull in food, it is apparent that the equivalence of these stimuli was primarily determined by the nature of the problem rather than the kind of previous learning. Klüver's monkeys had a 'direction' in their thought processes and the equivalent stimuli were influenced by it. Mental sets grow out of past experience and carry over into new situations, but 'directions' in thinking arise under the stress of a problem, and when they are new they cannot be traced to a previous situation. Because 'direction' and 'mental set' are determined differently it seems undesirable to reduce the concept of 'direction' to 'mental set.'

By distinguishing between equivalent stimuli which grow out of previous learning and those which are achieved in a problem situation, one can more easily determine where learning leaves off and reasoning begins in the solving of problems. Only the equivalent stimuli which are extensions of learning solve problems. The equivalent stimuli which are the product of reasoning are achieved, and cannot be given credit for producing the solution.

Once we recognize that the mechanism of equivalent stimuli can only solve some problems and make possible the formation of concepts under certain conditions, and that other problems and other concepts require a different mechanism, we can set about to determine the limitating conditions of the two different processes. By training people to react to different phases of a stimulus (e.g., the ceiling holds up the pillar and the pillar holds up the ceiling) the possible range of

solutions by equivalence is increased. On the other hand, emphasis on the fact that education is limited in that it cannot prepare a person for all types of situations and that specific problem situations can alter past experience will encourage the discovery of methods for decreasing the domination of past experience in behavior. Maier (10) for example, found that by removing certain inhibiting factors, the solving of problems involving reasoning was greatly increased. Since both types of functions have value in extending behavior possibilities, it is desirable that their analyses shall not be retarded by a confusion between them. The present investigation purports to point out the limitations of the mechanism of equivalent stimuli in the solving of creative problems, but at the same time recognizes its fundamental importance in the process of generalization and in the transfer of training. The experimental data are regarded as suggestive, and by no means do they settle the issues raised. It is hoped that more direct approaches to the problem may be made in the future, and that the indirect method used in this study may be extended and refined.

(Manuscript received March 23, 1945)

Bibliography

1. Duncker, K. A qualitative study of productive thinking. *J. genet. Psychol.*, 1926, **33**, 642–708.
2. Halstead, W. C. Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and non-equivalent stimuli. *Amer. J. Psychiat.*, 1940, **96**, 1263–1294.
3. Klüver, H. *Behavior mechanisms in monkeys.* Chicago: Univ. Chicago Press, 1933, pp. 387.
4. Maier, N. R. F. Reasoning and learning. *Psychol. Rev.*, 1931, **38**, 332–346.
5. Maier, N. R. F. Reasoning in rats and human beings. *Psychol. Rev.*, 1937, **44**, 365–378.
6. Maier, N. R. F. The specific processes constituting the learning function. *Psychol. Rev.*, 1939, **46**, 241–252.
7. Maier, N. R. F. The behavior mechanisms concerned with problem solving. *Psychol. Rev.*, 1940, **47**, 43–58.
8. Maier, N. R. F. Reasoning in humans. I. On direction. *J. comp. Psychol.*, 1930, **12**, 115–143.
9. Maier, N. R. F. Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *J. comp. Psychol.*, 1931, **12**, 181–194.
10. Maier, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, **24**, 144–155.
11. Maier, N. R. F. Qualitative differences in the learning of rats in a discrimination situation. *J. comp. Psychol.*, 1939, **27**, 289–331.
12. Maier, N. R. F. The effect of cortical injuries on equivalence reactions in rats. *J. comp. Psychol.*, 1941, **32**, 165–189.
13. Maier, N. R. F., & Schneirla, T. C. *Principles of animal psychology.* New York: McGraw-Hill, 1935, pp. 529.
14. Strauss, A. A., & Werner, H. Disorders of conceptual thinking in the brain-injured child. *J. nerv. ment. Dis.*, 1942, **96**, 153–172.
15. Wapner, S. The differential effects of cortical injury and retesting on equivalence reactions in the rat. *Psychol. Monogr.*, 1944, **57**, pp. 59.
16. Werner, H. Process and achievement, a basic problem of education and developmental psychology. *Harvard educ. Rev.*, 1937, May issue, 353–368.
17. Wertheimer, M. Über Schlussprozesse im Produktiven Denken. In: *Drei Abh. z. Gestalttheorie.* Erlangen: Philosophische Akademie, 1925, 164–184.

# THE EFFECTS OF ELECTROSHOCK CONVULSIONS ON DOUBLE ALTERNATION LEVER-PRESSING IN THE WHITE RAT

BY ELLIOTT McGINNIES* AND HAROLD SCHLOSBERG

*Brown University*

## Introduction

This paper reports an exploratory study of the effects of electroshock convulsions on double alternation behavior in the white rat. Only three animals were carried through to the end of the experiments, but the results were extremely suggestive. Since we are unable to continue the experiments at present, it seems best to publish our tentative findings now. Briefly, they indicate that although one or two electroshock convulsions will disrupt double alternation behavior, rats recover rapidly. Further, rats may become adapted to such convulsions, so that the habit is no longer disturbed by later members of a series.

In recent years cerebral electroshock has generally displaced other methods of inducing therapeutic convulsions. It is safer, and has the further advantage that the patient has no memory of the convulsion. It seems to be generally agreed that the series of electroshock convulsions is often valuable in the treatment of schizophrenia, cyclothymia, and even in some of the psychoneuroses. It seems clear that there is some amnesia for events before the convulsions, and that the patient is somewhat dazed for a period of an hour or so after the treatment (1). Some workers have suggested that the convulsions work through a psychogenic mechanism, but our experiments and those of Duncan (2) are in agreement with those who believe that there are changes in the brain tissue. This is not the place to review the extensive literature based on therapeutic use of the technique. We will be content to cite the few animal experiments that have been concerned with electroshock convulsions.

Page (6) subjected rats and cats to an extended series of electroshock convulsions. He used 18 milliamperes of 60 cycle current, at 600 volts, lasting two sec., to induce convulsions in cats. For rats, he reduced the current to 4.5 milliamperes, and passed it from nose to tail. The resulting seizures lasted about 30 sec., including tonic and clonic phases. After 50–100 seizures the rats became passive, inac-

* Now at Harvard University.

tive, and submissive. The change could be noticed after 10 or 15 convulsions when a control group was used for comparison.

Stainbrook and Lowenbach (**8**) studied maze behavior of rats after electroshock convulsions. Thirty animals were trained to select the correct exit from a simple two-choice water maze. They were then given a series of electroshock convulsions, produced by a potential of 80–85 volts, lasting 0.3 sec., applied through clips attached to the ears. The convulsions seemed about as severe as those obtained by Page (**6**). When the rats were placed in the maze 15 min. after the convulsion, the average swimming time was 6.6 sec., compared to 4 sec. during the original learning period. Stainbrook also mentioned the results of another experiment, in which rats were trained to run a maze and were then subjected to 30 daily convulsions. There was no impairment of maze relearning efficiency as measured by error scores, but the shocked rats were markedly worse than those of an unshocked group in terms of time scores. There was a marked fear reaction to handling in the experimental animals.

Duncan (**2**) studied the effect of electroshock convulsions on the Lashley Type III maze. After the rats had learned the maze, they were given 30 daily convulsions. Retention and relearning tests were started the day after the last shock. He used a shock similar to that reported by Stainbrook and Lowenbach (**8**). In addition to the shock convulsion group, there was a retention control group (30 day rest, no shocks) and a shock control group (30 daily shocks to the hind legs). He found a significant loss in retention and relearning when he compared the convulsive group to the two control groups, using trials, errors, or time as criteria. His experimental group did not show any fear reactions, although the shock control (leg shock) group was afraid of the experimenter. Since the shock control group showed no clear-cut loss in maze-running ability, while the experimental group showed such a loss, his results are opposed to the psychogenic theory.

The effect of electroshock convulsions upon the conditioned reflex in rats has been investigated by Kessler and Gellhorn (**5**). They conditioned rats to jump from one compartment to another at the sound of a bell, using a mild electric shock as the unconditioned stimulus. They then extinguished the response by failing to reënforce the bell with the shock. After 5 convulsive shocks, the conditioned response rose from the extinction level of between zero percent and 10 percent to 50 percent. When the shock was discontinued, the response fell back to the extinction level. A control group showed no spontaneous disinhibition of the response during the same length of time. The repeated success of this procedure showed conclusively that convul-

sions produced by electricity applied to the head induce a temporary recovery of an inhibited conditioned response.

Since the results in animal experiments were in disagreement among themselves and with the clinical reports, it seemed desirable to study the effects of electroshock on a more difficult task. One of us had shown that double alternation lever-pressing could be established in some rats and that it was a difficult habit (7). This learning situation has an advantage in that the rat can be given daily practice over a period of months without reaching a perfect score. A stable level of performance represents the capacity of that particular rat under those particular conditions. It seemed likely that changes in this level would be a sensitive indicator of any effects the electroshock convulsions might have.

The history of double alternation has been reviewed by Schlosberg and Katz (7), and need not be repeated here. It is sufficient to point out that Hunter and his students have found it to be too difficult a task for most rats. Of particular interest here is the experiment in which Hunter and Hall (4) showed that bilateral frontal lobe lesions destroyed the ability of three rats to run a double alternation maze. It seemed likely that cerebral electroshocks might have a similar effect. We used the lever-pressing situation instead of the double-alternation maze because it was more economical in time, since the rat can be left in the apparatus for long periods during the early stages of learning.

## Establishment of Double Alternation

*Apparatus.*—The apparatus used was the one mentioned in a footnote in the paper by Schlosberg and Katz (7). It consists of a Prestwood ratchet in which has been cut a series of N-shaped alleys forming a continuous double alternation maze. The rat moves a lever which slides a pin through the alleys and permits the ratchet to be advanced by a counterweight whenever the correct response is made. Reënforcement is given after any desired sequence of responses by food pellets placed in appropriate holes provided in the ratchet. These pellets fall to the rat's food dish through a tube entering the cage. In this manner, single right or single left responses may be rewarded, or food may be withheld until the completion of the double alternation sequence. The lever and food tube project into a cylindrical cage which restrains the rat.

In order to reduce friction and facilitate the early stages of learning, the ratchet should be milled carefully in a brass disc rather than in a strip of Prestwood. It was impractical to have this ratchet made during the war.

*Procedure.*—Four male rats, 3 to 4 months old, were used in the experiment. Training consisted, first, in establishing an association for the rat between the click of the apparatus and the presence of food in the dish. Wet mash was then spread on the lever to attract the rat's attention, and the pin was set so that the slightest pressure on the lever would advance the ratchet and send a food pellet into the feeding dish.

When all four rats were pressing the lever with a fair degree of regularity, an attempt was made to build up double alternation by placing them in the apparatus for four-hour periods and giving reënforcement only after two movements of the lever, either RR or LL.[1] When per-

[1] Upper case letters R and L indicate correct lever-pressures to right and left, respectively. Lower case letters l and r stand for pressures that were incorrect. Such errors did not advance the ratchet.

formance became stable at this level of training, reward was given only after completion of the complete double alternation sequence of RRLL. During this final period of training, the rats were allowed 40 runs (double alternations) daily, at the conclusion of which they were fed for 20 minutes on wet mash.

*Results.*—Only two rats reached the criterion of 80 percent correct double alternations in the daily recordings of 20 runs. Behavior early in training generally consisted of single alternation (RlRLrL) or partial double alternation (RlRLL). This latter habit has been previously noted by Schlosberg and Katz and is probably attributable to a tendency of the rat to persist in the response which is rewarded. The movements of the two proficient rats were clean and definite, and the sequence was uninterrupted.

The two rats which never mastered double alternation both appeared excited and over-active in the situation. Partial double alternation (RlRLL) became the most frequent response.

## Effect of Electroshock Convulsions

*Apparatus.*—Convulsions were produced by an 85 volt 60 cycle alternating current, lasting approximately 0.22 sec. and passed through the head from ear to ear. The current was obtained from the 120 volt lighting circuit and was cut down to 85 volts by a 100 ohm rheostat connected as a potentiometer. Voltage was checked on a voltmeter before each shock. The current was applied through small 'alligator' clips, padded with gauze, and soaked in a saline solution before they were clipped to the pinnae. Duration was controlled by manual operation of a toggle switch in series with the rheostat. This switch was thrown 'on' and 'off' in a definite rhythm by Mr. Duncan, who had had considerable practice in using the apparatus, but who knew nothing of the details of our experiment.[2]

*Shocking procedure and results.*—Since the experiment was dealing with an exploratory problem, no rigid procedure was followed. The initial step was to discover what change, if any, would occur in the behavior of one of the lever-pressing rats as a result of electrically induced convulsions. The observed decrement in double alternation performance suggested further investigation, the results of which, in turn, opened other leads. Since the results and procedures in the various stages are interdependent, it seems advisable to group them into three separate experiments. The method and results in each will be considered in turn, and relationships between them will be pointed out.

### *Experiment I:*

Rat number 4, who had consistently performed double alternation at the criterion level for about three weeks, was the first to be subjected to the electric shock. One electrode was fastened to each of the rat's ears, and the current was then applied for about 0.2 sec.

[2] When measured by a Springfield Timeclock, the duration rarely fell outside the limits of 0.20 and 0.25 sec.

A full convulsion, exhibiting both the tonic and clonic phases, resulted. It resembled in all respects the behavior previously described by Stainbrook and Lowenbach (**8**) and by Duncan (**2**). The rat's eyes closed, and the legs were first flexed under the body and then extended outward and backward as the animal went into full extensor rigidity. After 10 or 12 sec., the clonic phase began, in which the muscles of the body trembled spasmodically for a brief period. The rat was then replaced in the living cage, where it lay on its side breathing heavily for several minutes. Spontaneous movement generally did not occur until 5 or 10 min. after the end of the convulsion.

The shocking procedure with this rat was undertaken from 8 to 12 hours before running so as to insure ample time for recovery from the immediate effects of the shock. Rat 4 showed no apparent effects from the convulsion during the first test period. He made 17 correct double alternations in 20 trials. (Subsequently during the experiment, all rats were allowed 30 min. in which to complete 20 trials.) On the following day, the rat was again shocked in the same manner as before. When placed in the apparatus about 8 hours later, he failed to perform even one complete sequence. The following is a record of the behavior noted during this 30 minute period of observation. It is characteristic of the descriptions recorded for all the rats which exhibited performance breakdowns following shock.

Rat made some attempts to pull L (= Lever to Left) but appeared too enfeebled to do so. Prompting was of no avail. Rat showed lively interest in food dish when lever was moved or tapped. After seven min. of inactivity, rat pulled feebly and ineffectively at lever a few times. Was removed from apparatus for a short time and then replaced. Pulled R several times but was unable to pull L. Exhibited lethargy and some disequilibrium. Ate pellets when presented with them.

Complete recovery to the criterion level of performance did not take place until 10 days of training after the second shock. During this time, rat 2 was shocked through the rear legs on five successive days as a control for other possible effects of the shock. (See Table I.) This rat showed no decrement in performance; in fact, he improved slightly during the 5 day period. He did, however, show 'anxiety' when removed from his cage for shocking, and struggled to avoid the electrodes. Rats who were shocked through the head never displayed this type of resistance and were apparently amnesic to the experience. Contrary to Stainbrook's observations (**8**) but in agreement with those of Duncan (**2**), none of these rats developed a chronic fright state or behaved in a manner resembling an experimental neurosis.

It was now decided to investigate relearning time after a rest, on the assumption that the first few days of retraining were unprofitable because of the weakening effect of the convulsion. Rats 2 and 4 were

TABLE I

EXPERIMENT I

| Rat 2 | | | Rat 4 | | |
|---|---|---|---|---|---|
| June | Right | Wrong | June | Right | Wrong |
| 8 | 18 | 2 | 2 | 20 | 0 |
| 9 | 17 | 3 | 3 | 15 | 5 |
| 10 | 16 | 4 | 4 | 16 | 4 |
| 11 | 17 | 3 | 5 | 17 | 3 |
| 12 | 17 | 3 | 6 | 12 | 8 |
| 13 | 15 | 5 | 7 | 19 | 1 |
| 14 | 17 | 3 | 8 | 17 | 3 |
| 15* | 15 | 5 | 9† | 16 | 4 |
| 16* | 14 | 6 | 10† | 0 | 0 |
| 17* | 19 | 1 | 11 | 1 | 4 |
| 18* | 0 | 0 | 12 | 0 | 5 |
| 19* | 19 | 1 | 13 | 5 | 5 |
| 20 | 18 | 2 | 14 | 0 | 1 |
| 21† | 16 | 4 | 15 | 7 | 5 |
| 22† | 0 | 0 | 16 | 3 | 10 |
| | | | 17 | 12 | 8 |
| | | | 18 | 14 | 6 |
| | | | 19 | 14 | 6 |
| | | | 20 | 17 | 3 |
| | | | 21 | 19 | 1 |
| | | | 22† | 4 | 5 |
| | | | 23† | 0 | 0 |
| 20 days' rest | | | 20 days' rest | | |
| July | | | July | | |
| 14 | 3 | 1 | 14 | 0 | 0 |
| 15 | 3 | 4 | 15 | 12 | 6 |
| 16 | 3 | 7 | 16 | 14 | 6 |
| 17 | 5 | 4 | 17 | 9 | 11 |
| 18 | 5 | 11 | 18 | 16 | 4 |
| 19 | 10 | 6 | 19 | 17 | 3 |
| 20 | 17 | 3 | 20 | 16 | 4 |
| 21 | 20 | 0 | | | |
| 22 | 16 | 4 | | | |
| 23† | 0 | 0 | | | |
| 24 | 10 | 1 | | | |
| 25† | 0 | 0 | | | |
| 26 | 14 | 2 | | | |
| 27 | 0 | 0 | | | |
| 28 | 8 | 1 | | | |
| 29 | 15 | 2 | | | |
| 30 | 14 | 0 | | | |
| 31 | 14 | 1 | | | |
| Aug. | | | | | |
| 1 | 18 | 2 | | | |
| 2 | 17 | 3 | | | |

* indicates day after shock through rear legs.
† indicates day after cerebral shock.
The daily performance of rats 2 and 4 before and after electroshock convulsions. The entries give the numbers of correct and incorrect double alternation sequences during 20 responses (or 30 min., if responses were infrequent.)

used as controls on each other. Following recovery from the first two shocks, rat 4 was shocked twice again, breaking his performance down to zero. (See Fig. 1.) The same thing was done to rat 2 who had never been shocked before. Both rats were then given a 20 day rest

without any practice. At the end of this time, they were retrained to criterion. Rat 4 required five days, and rat 2 took seven days. It was found that relearning time for rat 4 was shorter by five days than the previous recovery period when constant practice had followed the breakdown. Rat 2 was now shocked again until he last the habit, two shocks being sufficient. Unfortunately the two shocks were not on successive days. The habit fell to zero after one convulsion, so the second shock was omitted. When the habit returned

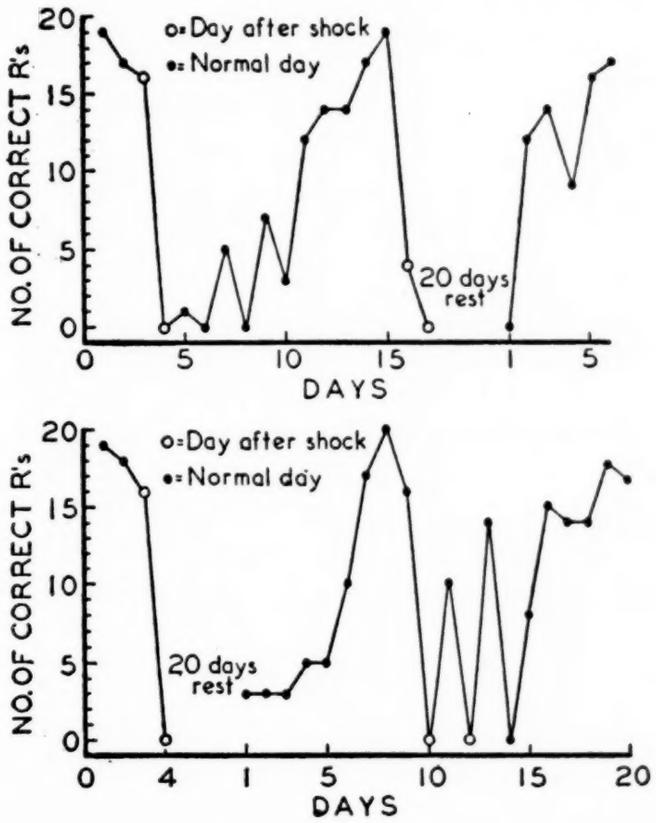



FIG. 1. Effect of rest vs. retraining on recovery from effects of electroshock convulsion. (See Experiment I.) Upper graph represents rat 4; lower graph, rat 2.

on the following day, another shock was given. The recovery from these two shocks was quite irregular, making it dangerous to decide when recovery was complete.

The behavior of these animals in the lever-pressing situation following the rest period of 20 days was qualitatively different from their behavior immediately following shock. Both animals, for one thing, showed less irregularity in performance. That is, they made steady improvement up to criterion level (with one drop in the case of rat 4), whereas in relearning immediately after shock they had both 'good' and 'bad' days until criterion was reached. Movements of the lever following rest were made with certainty rather than with hesitancy and insufficient force, as when following a convulsion. Errors, how-

ever, were of the same order as those caused by convulsions. They consisted of a regression to the earlier responses of RlRLL, with occasional single alternations, RlRLrL.

The results of this experiment suggest that electroshock convulsions impair the performance of a rat for several days. This impairment does not seem to be a permanent loss of the habit. Rats recover the habit at least as rapidly after a 20 day vacation following the shocks as they do if training is not interrupted. The results are somewhat complicated by a 'practice effect' (Experiment II), but the behavior of the rat strongly suggests a failure of performance rather than of retention.

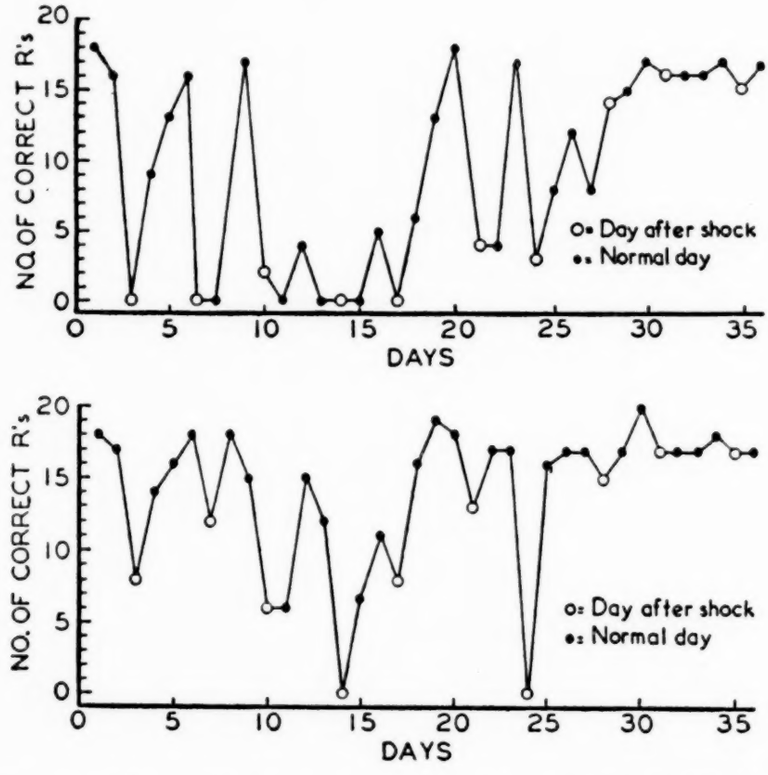



FIG. 2. Effects of bi-weekly electroshock convulsion on the double alternation habit. (See Experiment II.) Upper graph represents rat 4; lower graph, rat 2.

*Experiment II:*

The next question which arose following the initial findings was whether a series of shocks might have a cumulative effect, so as to cause a permanent or long-lasting breakdown of the double alternation habit. Rats 2 and 4 were run at criterion for several days to reestablish the habit firmly and were then placed on a schedule of two shocks weekly. Rat 2 was shocked at night and run in the morning on the following day. Rat 4 was shocked in the morning and run at night on the same day. Shocks were administered on Tuesdays and Fridays, and a total of 10 were given each rat.

Both rats reacted in the same general fashion to this series of shocks, although they differed slightly in details. (See Fig. 2.) During the first 10 days of the schedule, there was a tendency for the performance level of rat 4 to recover rapidly between shocks. Then, for a period of approximately 15 days, performance remained consistently at a low level. During the last 11 days, the performance level rose steadily until it finally stabilized above criterion in spite of shocking. Rat 2 presented a slightly different picture. For the first 10 days, his performance stayed at a relatively high level, despite shocking, and showed only a moderate drop on the days following convulsions. Then, for about the middle 15 days, performance was highly irregular. It dropped to zero on the days following two of the shocks and then recovered rapidly before the next scheduled shock. As in the case of rat 4, the performance level reached a point at or above criterion during the last 11 days and was resistant to the convulsions.

In general, it may be said that both rats showed an initial susceptibility to the shocks but recovered rapidly from them. This was followed by a period of uniformly poor performance in one rat and highly variable performance in the other rat, possible due to a cumulative effect of the convulsions. Finally, both rats developed a 'resistance' to the effects of the convulsions and performed consistently at criterion or better, even on the days following shock. The convulsions, however, continued uniformly throughout the series of shocks.

*Experiment III:*

An attempt was made to use the relearning-control technique of Experiment I on rats 3 and 5, which had never completely mastered double alternation. The highest score made by either rat was 50 percent correct responses, and the mean was about 20 percent for rat 3 and 25 percent for rat 5. Rat 3 refused to press the lever nine hours after his first convulsion but behaved normally in other respects. After a rest of three days, he recovered to normal in two days of training, the criterion being 20 responses in 30 min. regardless of errors. When shocked again and retained immediately, he took five days to relearn. This indicated that the first three days following shock constituted a period of spontaneous recovery which was unaffected by practice. (See Experiment I.)

Rat 5 broke down after 4 shocks and recovered, with constant training, in five days. It was now planned to repeat the shock and to rest him for the first three days, as had been done with rat 3. However, he reacted surprisingly to the shock by improving from 35 per-

## TABLE II

### Experiment III

The daily performance of rats 3 and 5 before and after electroshock convulsions. The entries give the numbers of correct and incorrect double alternation sequences during 20 responses (or 30 min., if responses were infrequent).

| Rat 3 | | |
|---|---|---|
| July | Right | Wrong |
| 21 | 7 | 13 |
| 22 | 6 | 14 |
| 23 | 5 | 15 |
| 24 | 0 | 20 |
| 25* | 0 | 0 |
| 2 days' rest | | |
| 28 | 1 | 15 |
| 29 | 2 | 18 |
| 30 | 1 | 19 |
| 31* | 0 | 0 |
| August | | |
| 1 | 0 | 2 |
| 2 | 0 | 3 |
| 3 | 2 | 16 |
| 4 | 1 | 19 |
| 5 | (not run) | |
| 6 | (not run) | |
| 7 | 1 | 19 |
| 8 | 0 | 20 |
| 9* | 1 | 19 |
| 10* | 0 | 1 |
| 11* | rat paralyzed | |

| Rat 5 | | |
|---|---|---|
| July | Right | Wrong |
| 28 | 5 | 12 |
| 29 | 3 | 17 |
| 30 | 5 | 15 |
| 31 | 4 | 16 |
| August | | |
| 1* | 1 | 19 |
| 2* | 6 | 14 |
| 3* | 7 | 5 |
| 4* | 0 | 2 |
| 5 | 0 | 0 |
| 6 | 0 | 0 |
| 7 | 6 | 7 |
| 8 | 3 | 9 |
| 9 | 6 | 14 |
| 10 | 7 | 13 |
| 11* | 6 | 11 |
| 12* | 8 | 8 |
| 13* | 13 | 5 |
| 14* | 9 | 1 |
| 15 | 11 | 8 |
| 16 | 13 | 7 |
| 17 | 5 | 12 |
| 18 | 13 | 7 |
| 19* | 10 | 10 |
| 20* | 13 | 7 |
| 21* | 15 | 5 |
| 22* | 12 | 8 |
| 23* | 11 | 9 |
| 24* | 10 | 10 |
| 25* | 10 | 10 |
| 26* | 12 | 8 |
| 27* | 16 | 4 |
| 28* | 13 | 7 |
| 29* | 10 | 10 |
| 30* | 7 | 13 |
| 31* | 7 | 13 |
| Sept. | | |
| 1* | 10 | 10 |
| 2* | 9 | 11 |
| 3* | 9 | 11 |
| 4* | 10 | 10 |
| 5* | 8 | 12 |
| 6* | 11 | 9 |
| 7 | 12 | 8 |
| 8 | 4 | 16 |
| 9 | 5 | 15 |

* indicates day after cerebral shock.

cent correct responses on the day preceding the first convulsion to 45 percent correct responses on the day following four successive convulsions. Shocks were then discontinued to study the effect, and the rat's performance rose to an average of 52.5 percent correct responses for four days without shock. With the shocks again administered daily, correct double alternations made by the animal rose to an average for the next four days of 62.5 percent. Shocks were now continued until the animal had had 19 daily convulsions and was still performing on the lever with an average accuracy for the total period of 55 percent. The shocks were again discontinued, and the average number of correct responses for the four days following fell to 40 percent. (See Table II.)

Attempts to discover the effects of daily convulsions upon rat 3 were terminated after the animal was paralyzed from the cervical region downward following the third shock. The animal died a week later, presumably due to a spinal hemorrhage caused by the convulsion.

## Discussion

Perhaps the most often recurring question in connection with the various shock therapies concerns their effect upon memory. Investigators are currently in disagreement as to whether serious memory disturbance results from shock convulsions. In our experiment, the question at once arises as to whether the results indicate a loss of recent memory in rats following shock, or whether some other explanation of the behavior decrements is possible. The fact that results with the different animals were somewhat diverse makes it difficult to formulate a single, consistent theory to explain all the data.

For example, on the basis of a single hypothesis, how can we explain the performance decrements shown by rats 2 and 4 and, at the same time, account for the increment in performance of rat 5? One possible explanation is as follows: The effect of the convulsions was such as to cause a general physical weakening of the animal, i.e., loss of muscle tonus, lowering of activity level. In rats 2 and 4, this enfeeblement decreased their motivation to a point where performance in the apparatus was severely impaired. Rat 5, which had always been excited and over-active, was 'slowed down' by the convulsions and consequently was less prone to commit errors. Furthermore, rat 5 was larger and heavier than the other animals, and it is reasonable to suppose that he was therefore more resistant to the convulsions.

An alternative explanation might also be offered. Since the presence of cerebral anoxia during convulsions has been demonstrated (3), it may be that anoxia in the rats depressed cerebral func-

tion, causing a temporary loss of retention. Performance decrements might even be the joint result of a drop in activity and of the temporary eclipse of the cortical process controlling the habit.

Presence of a spontaneous recovery period observed in rats 2, 3, and 4 would indicate that no permanent structural injury was occasioned by the convulsion. It is difficult, moreover, to see how the performance levels of rats 2 and 4 could have reached a final high plateau after 10 convulsions if severe cerebral injury had resulted.

Observational data on the rats who suffered a breakdown in performance following one or more convulsions militate against the theory of recent memory loss as an explanation for these results. Rat 4, on several occasions in post-shock trials, went through the motions of moving the lever but did not exert enough force to be effective. Moreover, the low scores made by all of the rats during the relearning period were often due to partial completion of the trials rather than to errors. The fact that the animals did occasionally revert to an earlier form of response, namely, RlRLL, suggests that the mechanism controlling the response may have been in some way affected by the convulsion. But this same error occurred even during normal performance, so it is not clear evidence for memory loss.

## Summary and Conclusions

Double alternation lever-pressing to a criterion of 80 percent accuracy, or better, was established in two albino rats. Two other rats learned to make up to 50 percent correct responses, with partial double alternation (RlRLL) constituting the most frequent errors.

Convulsions, induced by passage of an electric current through the head, caused a breakdown of the habit in all of the rats during the initial stages of the experiment. Subsequent variations in the procedure showed three types of reaction to the shock:

1. The rats (2 and 4) who had learned to criterion, each exhibited breakdown of performance following each of two pairs of cerebral shocks. In each rat, one pair of shocks was followed by a rest period of 20 days, and the other pair was succeeded by immediate retraining. The results suggest that electroshock convulsions are followed by a short period during which the rat cannot profit from retraining.

2. Rats 2 and 4 were given a series of convulsions, induced twice weekly, and accompanied by daily practice in the apparatus. The reaction to the initial shocks was sudden breakdown followed by rapid recovery. This was succeeded by a period of variable or generally poor performance. Finally, both animals again reached criterion and remained there despite the continuance of bi-weekly convulsions.

3. Rat 5 which had never reached criterion was placed on a schedule of daily shocks. He improved steadily during the period of convulsions and, following 19 consecutive shocks, performed better than he had for any comparable period prior to the shocks.

No evidence of experimentally produced neurosis was observed in any of the animals. All were apparently amnesic to the convulsive incident.

The theories offered to explain the behavior were: first, that the convulsions decreased activity level, thereby impairing performance; and second, that cerebral anoxia induced by the convulsion caused a temporary loss of retention.

Generalization of these results to the effects of electroshock convulsions in human patients must await further experimentation. Preliminary evidence indicates that the convulsions have no permanent debilitative effects in some individuals, aside from actual bodily injury caused by the violence of the convulsion.

(Manuscript received March 5, 1945)

## References

1. Cook, L. C. Convulsion therapy. *J. ment. Sci.*, 1944, **90**, 435–464.
2. Duncan, C. P. The effect of electroshock convulsions on the maze habit in the rat. *J. exp. Psychol.*, 1945, **35**, 267–278.
3. Himwich, H., Bowman, K., Wortis, J., & Fazekas, J. Metabolism of the brain during insulin and metrazol treatment of schizophrenia. *J. Amer. med. Ass.*, 1939, **112**, 1532.
4. Hunter, W. S., & Hall, B. E. Double alternation behavior of the white rat in a spatial maze. *J. comp. Psychol.*, 1941, **32**, 253–266.
5. Kessler, M., & Gellhorn, E. The effect of electrically and chemically induced convulsions on conditioned reflexes. *Amer. J. Psychiat.*, 1943, **99**, 687–691.
6. Page, J. D. Studies in electrically induced convulsions in animals. *J. comp. Psychol.*, 1941, **31**, 181–194.
7. Schlosberg, H., & Katz, A. Double alternation lever-pressing in the white rat. *Amer. J. Psychol.*, 1943, **56**, 274–282.
8. Stainbrook, E. J., & Lowenbach, H. The reorientation and maze behavior of the rat after noise-fright and electroshock convulsions. *J. comp. Psychol.*, 1942, **34**, 293–299.

# TONAL RANGE AND VOLUME LEVEL PREFERENCES OF BROADCAST LISTENERS

BY PHILIP EISENBERG AND HOWARD A. CHINN

*Columbia Broadcasting System, Inc.*

## Problem

The engineer is frequently confronted by problems in sound reproduction which take him into the realm of psychology. He soon finds that his task involves more than physical measurements and readings on a dial. He must know how listeners react to auditory stimuli.

An excellent illustration of this point is the problem of tonal range preference. Engineers have reached the stage in the development of techniques where they can reproduce sound through transcriptions and over the radio at a wide enough tonal range to capture all the overtones and harmonics of the highest and lowest pitched instruments. In their research, engineers have determined the frequency range of the human ear (**2**), the manner in which this characteristic varies with age (**8**), the discernibility of changes in frequency range (**4**), the requirements of an ideal program transmission system (**3**) and the audible frequency ranges of musical instruments and of speech (**7**). But they did not bother to discover whether listeners actually liked to hear sound reproduced over such wonderfully developed equipment. It was taken for granted that the listener would like it.

It is the purpose of this study to discover what the tonal range preferences of listeners actually are. Since volume level is a very important element in sound, volume level preferences were also investigated. One further question was raised: to what extent is the listener influenced by the judgment of his own ears as compared with the influence of prestige suggestion?

## Procedure

### *Basic plan*

The study consisted of six experiments, in each of which the participants were presented with 10 passages of speech and 10 passages of music. During each passage the form of reproduction was alternated between two different tonal ranges or two different volume levels. For some passages, volume was kept constant while tonal range varied; for others, the tonal range was kept constant while intensity was varied; and in a few cases, both factors were changed together.

For each passage, the participants were asked to select from the two methods of presentation, the one they found more *pleasant* to listen to. They were asked to react to broadcasts in the same way as they would at home. The overwhelming majority of listeners in all experiments indicated that they would not have made different choices if they were listening at home.

It was found necessary to conduct several types of experiments, rather than one, since preliminary studies indicated that the nature of the tests were such that fatigue or boredom might be influencing factors after the *S*s had made judgments for more than 20 or 25 passages. The experiments were designed as follows:

*Experiment 1* was a basic experiment primarily for the purpose of determining tonal range preferences of a cross-section of listeners for male speech and classical music.

*Experiment 2* involved the presentation of types of program material and of performers not covered by the first experiment. The program content consisted of female speech, mixed dramatic speech, piano and popular music.

*Experiments 3 and 4* were designed to test the tonal range preferences of two special groups: professional musicians and FM listeners respectively. The group of professional musicians was included in the studies to determine whether past experience with wide tonal range influenced preferences. Both groups were used to discover whether any preferences found with a cross-section of listeners would be similar for special groups; that is, to test the universality of preferences. The plan of these experiments followed that of the first experiment exactly and employed the same program material.

*Experiment 5* was undertaken to discover whether recordings may have influenced the results. In all other experiments electrical transcriptions were used. The plan of this experiment was the same as for the other tonal range experiments except that a regular broadcast program, transmitted by wire directly from the originating studio, was used instead of recordings.

*Experiment 6* was primarily for the purpose of determining volume level preferences. The same male speech and classical music selections used in Experiments 1, 3 and 4 were presented in these tests. The list of experiments is summarized in Table I.

TABLE I

List of Experiments

| Experiment | Basic Type | Program Material | Subjects | No. of Subjects |
|---|---|---|---|---|
| 1 | Tonal Range | Voice and Classical Music (recorded) | Broadcast Listeners | 105 |
| 2 | Tonal Range | Female Speech, Mixed Dramatic Speech, Piano, Popular Music (recorded) | Broadcast Listeners | 123 |
| 3 | Tonal Range | Voice and Classical Music (recorded) | Professional Musicians | 20 |
| 4 | Tonal Range | Voice and Classical Music (recorded) | FM Listeners | 96 |
| 5 | Tonal Range | Light Classical and Popular Music (live) | Broadcast Listeners | 43 |
| 6 | Volume Level | Voice and Classical Music (recorded) | Broadcast Listeners | 111 |
| | | | Total | 498 |

*Subjects*

The main groups of *S*s, all adults, were secured by means of spot announcements broadcast over the CBS key station, WABC, in New York City. The professional musicians who took part in the tests were CBS staff performers with many years of training and experience. The FM listeners were obtained from a list of FM set owners that had been compiled in connection with another study.

A total of 29 listener sessions were conducted for the various experiments. In no case were the same *S*s used for more than one experiment. A total of 498 listeners took part in the tests as indicated in Table I. Altogether, over 10,000 individual preferences were obtained.

*Environment*

In order to simulate living room conditions, a small studio with a low ceiling was used for the study. It was 22 ft. wide, 30 ft. long, and 8½ ft. high. It was provided with a rug, piano, armchairs and other furnishings to give it a living room atmosphere.

The loudspeaker used to reproduce the voice and program material was located at the front of the room. It was hidden behind a very light scrim curtain. Located at the other end of the room, was a sound proof control booth in which the transcriptions were played. It also was hidden from the view of the listeners by a screen.

Measurements indicated that the sound intensity in all parts of the studio was essentially alike, since the loudspeaker had a wide angle of coverage. As an additional precaution, however, all listeners were seated toward the front of the room and as near its center line as feasible.

*Equipment*

Except for one series of tests made with live talent, all voice and musical selections were reproduced from especially recorded 'masters' cut on cellulose-nitrate coated discs. The records were made by the Columbia Recording Corporation and employed the standard electrical tran-

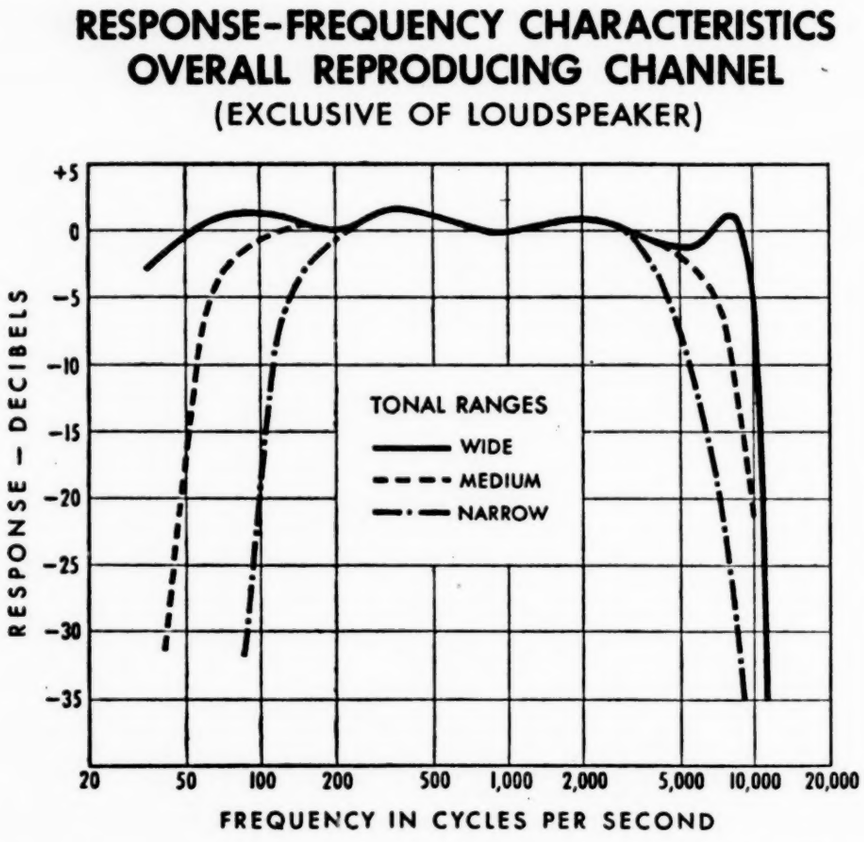



FIG. 1. The frequency limits of the three tonal ranges for the entire channel, including the recording and reproducing equipment, but excluding the loudspeaker

scription recording characteristic (6). This source of program material was used to insure absolute uniformity in the material presented at each session. Furthermore, original master recordings were used because of the extremely low surface noise level that this type of recording affords. A new cut was used for each session to avoid any possibility of the quality being impaired by repeated playings.

The overall response of the reproducing system was essentially uniform from 40 to 10,000 cycles. The background noise, even during the reproduction of the records, was negligible and not detectable by the majority of the listeners. The measured distortion of the electrical portion of the system was extremely low throughout the frequency range. Although facilities were not available to determine, quantitatively, the non-linear distortion introduced by the loudspeaker, trained observers, critically listening to the system, agreed that no distortion of this type was detectable. This observation was consistent with the manufacturer's claims for the loudspeaker and the measured distortion of the electrical portion of the system.

The loudspeaker was a dual unit employing a folded horn for the low frequencies and a multi-cellular horn for the highs; the coverage of the latter was uniform over a 70 degree angle in

the horizontal plane and a 35 degree angle in the vertical plane. The loudspeaker was installed high enough above the floor so that listeners in the front row did not block the sound for listeners in succeeding rows.

The system was equipped with filters to provide three tonal ranges, nominally designated as narrow, medium and wide. The overall response frequency characteristics of the system (including the recording and reproducing equipment but exclusive of the loudspeaker) is illustrated in Fig. 1. The transmission bandwidths were chosen, in accordance with established CBS practice (1), so as to maintain a balance between the high and the low frequencies. Furthermore, the band-pass filters had the characteristics shown, rather than very sharp cut-offs, since these are representative of the conditions that generally prevail in radio and recording equipment.

*Program material*

In Experiments 1, 3, 4 and 6 two classical musical selections (Music A and B) and two male speech selections (Speech A and B) were used (see Table II). In order to ascertain the effect

TABLE II

LIST OF PROGRAM MATERIAL

| Selection | Title | Author | Performer | Used in Experiment |
|---|---|---|---|---|
| | | *Music selections* | | |
| A | Carmen | Bizet | Orchestra, 40 pcs. | 1, 3, 4, 6 |
| B | Preludia | Jarnefelt | Orchestra, 40 pcs. | 1, 3, 4, 6 |
| C | Moon Love | McDavid | Pianist | 2 |
| D | Long Ago and Far Away | Kern | Orchestra, 16 pcs. | 2 |
| E | An Hour Never Passes | Kennedy | Soprano and Orchestra† | 5 |
| F | If I Knew Then | Jurgens | Soprano and Orchestra† | 5 |
| G | I'll Be Seeing You | Kahal | Chorus* | 5 |
| H | Sweet and Lovely | Arnheim, Tobias, Lemas | Tenor and Orchestra† | 5 |
| I | Amor | Ruiz | Soprano and Orchestra† | 5 |
| J | Pretty Kitty Blue Eyes | Curtis | Tenor, Chorus,* and Orch.† | 5 |
| K | Long Ago and Far Away | Kern | Soprano and Orchestra† | 5 |
| L | Let Me Call You Sweetheart | Friedman | Chorus and Orchestra† | 5 |
| M | A Kiss to Remember | Silver | Soprano, Chorus,* Orch.† | 5 |
| N | My Jesus, I Love Thee | Gordon | Soprano, Chorus,* Orch.† | 5 |
| | | *Speech selections* | | |
| A | Script 'Untitled' | Corwin | Tenor Speech | 1, 3, 4, 6 |
| B | Script 'Untitled' | Corwin | Tenor Speech | 1, 3, 4, 6 |
| C | Script 'Untitled' | Corwin | Soprano Speech | 2 |
| D | Mrs. Miniver | Bixby | Soprano and Baritone Speech | 2 |

* 14 female voices. † 29 pieces.

of program content upon the choices, Experiment 2 was performed with other types of program material. The supplemental pieces were a piano solo covering a wide tonal range (Music C), a popular song played by a 16 piece orchestra (Music D), female speech (Speech C) and speech with mixed voices and sound effects (Speech D).

For the tests undertaken with live talent (Experiment 5), a regular CBS broadcast program, 'The American Melody Hour,' was transmitted directly from the originating studio to the experimental studio with no degradation in fidelity. The selections were of a light popular variety, played by a 29-piece orchestra, with vocals by a 14-voice female chorous, a tenor and two sopranos. The selections are detailed in Table II, Music E to N inclusive. During this experiment the listeners heard only the musical selections; all spoken material was deleted.

*Experimental procedure*

The pattern followed in all listener sessions was alike. The *Ss* were told that, during the course of the experiment, they were to hear music and speech presented in several different ways. The listeners were advised that there were no right or wrong answers, and that they were to indicate the method of presentation they found more pleasant and personally liked. In addition, the blanks upon which the *Ss* recorded their choices provided means for indicating if both conditions were liked about equally well or neither was liked. Furthermore, if the listeners had a strong preference they could so indicate. The instructions were recorded and played to the listeners at the beginning of each session.

The conditions of presentation were identified for the listeners by means of a pair of signal lights, numbered 1 and 2, which were synchronized with the changes in type of presentation. The lights were mounted vertically, one over the other. It was found that the listeners were not influenced in any way by the position of the lights in their choices.

Each test passage was exactly one min. in length, and during this period the paired conditions were alternately presented, each time for a 10-sec. interval. Thus, in a one-min. period, the listener had an opportunity to compare each type of presentation three times.

In the experiments, as previously pointed out, three tonal ranges and three degrees of volume level were used in various combinations. For convenience, these are nominally referred to as narrow, medium and wide tonal ranges; and low, moderate and high volume levels. The actual tonal ranges are shown in Fig. 1. The volume levels were: low, 50 decibels above the acoustical reference level; moderate, 60 decibels; and high, 70 decibels.

These particular limits were chosen because the narrow tonal range is representative of the performance of existing console type radio sets. The more popular table models have even narrower tonal ranges (5). On the other hand, the wide tonal range is believed to be well above the capabilities of all radio sets with the exception of a few laboratory type installations. The intermediate, or medium tonal range was chosen since, based on liminal units, it lies exactly half way between narrow and wide tonal ranges (4).

In the tonal range Experiments 1 through 5, five basic pairs were used for comparison as shown in Table III. In each of the first three comparisons, the tonal ranges were varied and the

TABLE III

BASIC PAIRS FOR COMPARISON TESTS

| Pair No. | Conditions |
|---|---|
| | *Experiments 1, 2, 3, 4 and 5* |
| 1 | Narrow vs. medium tonal range, both at moderate volume |
| 2 | Medium vs. wide tonal range, both at moderate volume |
| 3 | Narrow vs. wide tonal range, both at moderate volume |
| 4 | Wide range at moderate volume vs. narrow range at high volume |
| 5 | Wide range at high volume vs. narrow range at moderate volume |
| | *Experiment 6* |
| 6 | Low vs. moderate volume, both wide tonal range |
| 7 | Moderate vs. high volume, both wide tonal range |
| 8 | Low vs. moderate volume, both narrow tonal range |
| 9 | Moderate vs. high volume, both narrow tonal range |
| 10 | Wide range at low volume vs. narrow range at moderate volume |

volume kept constant at a moderate (60 db) level. In the last two pairs, both tonal range and volume were varied.

In the volume level tests, Experiment 6, five other basic pairs were used as shown in Table III. In each of the first four comparisons, the volume level was varied and the tonal range kept constant; for the first two, the tonal range is wide, and for the last two, narrow. In the last comparison, both tonal range and volume are varied.

In each experiment the five comparisons were repeated for each of the four types of program content (two music and two speech) making a total of 20 paired comparisons for each experiment, except for Experiment 5 in which only 15 pairs were presented. The order of presentation was varied and the light positions were reversed from the first music and speech selections to the second so that order and position effects might be eliminated. For example, condition 1 for

TABLE IV

TONAL RANGE PREFERENCES

| | Tonal Range | Experiment 1 (Cross-Section) | | Experiment 2 (Cross-Section) | | | | Experiment 3 (Musicians) | | Experiment 4 (FM Listeners) | | Experiment 5 (Live Talent) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Classical Music (A & B) | Male Speech (A & B) | Piano Music (C) | Pop. Music (D) | Female Speech (C) | Mixed Speech (D) | Classical Music (A & B) | Male Speech (A & B) | Classical Music (A & B) | Male Speech (A & B) | Light Classical Music (E to H and K to M) |
| Pair 1 | Narrow Range | 38% | 25% | 20% | 26% | 33% | 34% | 28% | 10% | 28% | 31% | 26% |
| | Medium Range | 19 | 52 | 25 | 33 | 46 | 34 | 20 | 62 | 31 | 55 | 21 |
| | No Preference | 43 | 23 | 55 | 41 | 21 | 32 | 52 | 28 | 41 | 14 | 53 |
| Pair 2 | Medium Range | 67% | 55% | 24% | 22% | 26% | 64% | 83% | 48% | 61% | 63% | 70% |
| | Wide Range | 12 | 21 | 28 | 39 | 30 | 15 | 7 | 25 | 16 | 14 | 11 |
| | No Preference | 21 | 24 | 48 | 39 | 44 | 21 | 10 | 27 | 23 | 23 | 19 |
| Pair 3 | Narrow Range | 58% | 48% | 30% | 34% | 29% | 45% | 73% | 40% | 59% | 48% | 40% |
| | Wide Range | 15 | 24 | 24 | 33 | 34 | 23 | 5 | 48 | 28 | 36 | 26 |
| | No Preference | 27 | 28 | 46 | 33 | 37 | 32 | 22 | 12 | 13 | 16 | 34 |
| | Size of Sample | 105 persons | | 123 persons | | | | 20 persons | | 96 persons | | 43 persons |

Music A became condition 2 for Music B. Each pair with the same content was repeated to obtain an estimate of reliability.

At the end of the tests, in Experiments 1 through 4, the listeners were informed that they had been making choices between material presented at low and high fidelity. A brief description of low and high fidelity was read. They then heard two more passages, one with a wide tonal range and the other with a narrow range, with a knowledge of which signal light corresponded to each. Since the term 'high fidelity' carries with it prestige that the term 'low fidelity' does not, this test, in a sense, compared the influence of suggestion with the judgment of the listener's ear.

## Results

### *Listener tonal range preferences when not informed of conditions*

In each of the Experiments 1 through 5, there were three pairs of tonal range comparisons: (1) narrow vs. medium, (2) medium vs. wide, and (3) narrow vs. wide (see Table III). In all cases the peak sound intensity was held constant at a 60 db acoustical level. Table IV presents the choices made for music and for speech for all the tonal range experiments. Since the results for Music A and Music B were very much the same, they were combined in the table. The same is true for Speech A and Speech B.

In *Experiment 1*, a cross-section of listeners were the *S*s and the same male voice and classical music passages were reproduced for all the different tonal ranges. Here it is evident that when the narrow tonal range is compared with the medium range (Pair 1), there is a preference for narrow with music and medium with speech. However, in the comparisons between medium and wide (Pair 2), and between narrow and wide (Pair 3), the preferences are very markedly in favor of the narrower band for both classical music and male speech. These preferences are graphically illustrated in Fig. 2.

In *Experiment 2*, a variety of program content, not covered by the preceding experiment, was presented to a cross-section of listeners. In this experiment the preferences were not marked in either direction, that is, there was no distinct choice for any of the pairs in most cases. In addition, an unusually large number of choices indicated no preference. Only with mixed speech is there a marked preference for the medium over the wide range, and for the narrow over the wide range. Whereas, in the first experiments, the preference fell somewhere between a narrow and medium tonal range, in this experiment the preference seemed to lie in the medium range. Apparently the preference varies somewhat with type of program content. But in any event, in no instance was there a clear preference for a wide band.

The *S*s for *Experiment 3* were professional musicians with considerable musical training and experience. The FM listeners who were *S*s of *Experiment 4* all owned an above average console model FM set for at least one year and most of them used it 'frequently' or

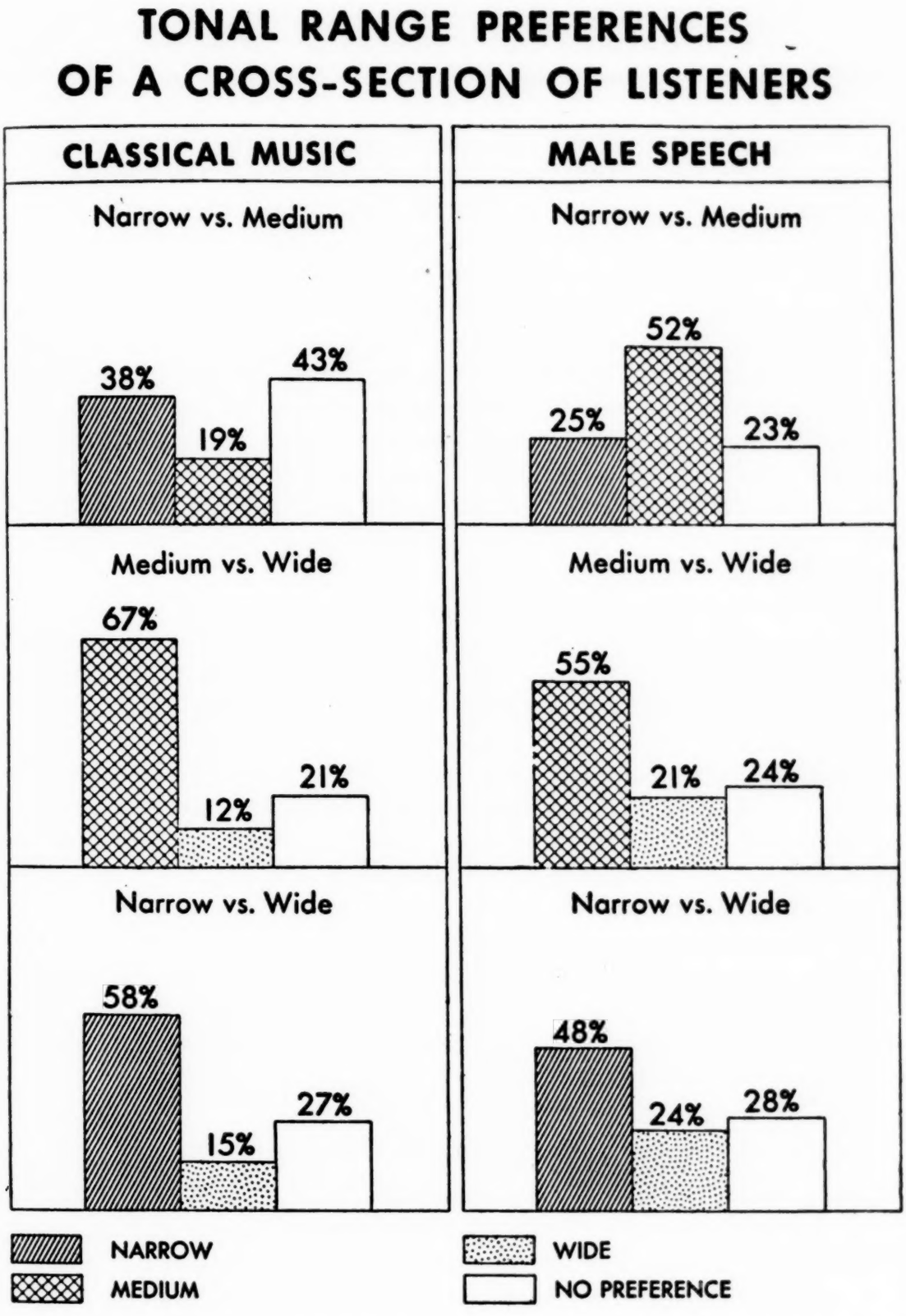



FIG. 2. The tonal range preferences of a cross-section of listeners for classical music and male speech

'regularly' according to their own statements. In addition, this group had more education and more musical training and experience than the cross-sections of listeners who acted as Ss in Experiments 1, 2 and 5.

The results for these two special groups with male speech and classical music were substantially the same as for the cross-section of listeners in Experiment 1; that is, the preference was for a tonal range somewhere between narrow and medium. Furthermore, the amount of FM listening did not seem to affect tonal range preference, since the choices of the 'regular' listeners were similar to the choices of the 'occasional' listeners. Thus, it would appear that the preference for a narrow and medium tonal range is wide spread, since such choices were made even by two groups who had considerably more musical experience than ordinary groups. The results for the professional musicians also tend to disprove any hypothesis that a narrow or medium tonal range preference is the result of limited experience with a wide range. Fig. 3 illustrates the tonal range preferences of the professional musicians.

In *Experiment 5*, a cross-section of listeners expressed their preferences based on tests involving live talent, playing and singing light classical and popular music. Here, when a narrow and a medium band were compared, most of the judgments were equal and no marked preference for either is evident. However, as in the case of preceding experiments, there was a very marked preference for medium over wide, and a consistent preference for narrow over wide tonal ranges. It is seen, therefore, that the results obtained in the preceding experiments could not have been influenced by the use of electrical transcriptions.

In *summary*, the results of the above experiments indicate that listeners predominantly prefer a tonal range somewhere between a narrow and medium band. Although the preferences vary somewhat for different types of music and speech, the overwhelming majority *never* choose a wide band.

*Listener tonal range preferences when informed of conditions*

Even when the listener has been informed as to which condition is 'low fidelity' (narrow tonal range) and which is 'high fidelity' (wide tonal range), he still chooses low fidelity. In the tests of *Experiments 1, 2 and 3*, in which the listeners are informed of conditions, the preference for the narrow band is as great as when listeners are not informed. (See Table V.) However, there is a slight increase in the number who preferred the wide band, but this is at the expense of the equal judgments. In all probability, the increased preference

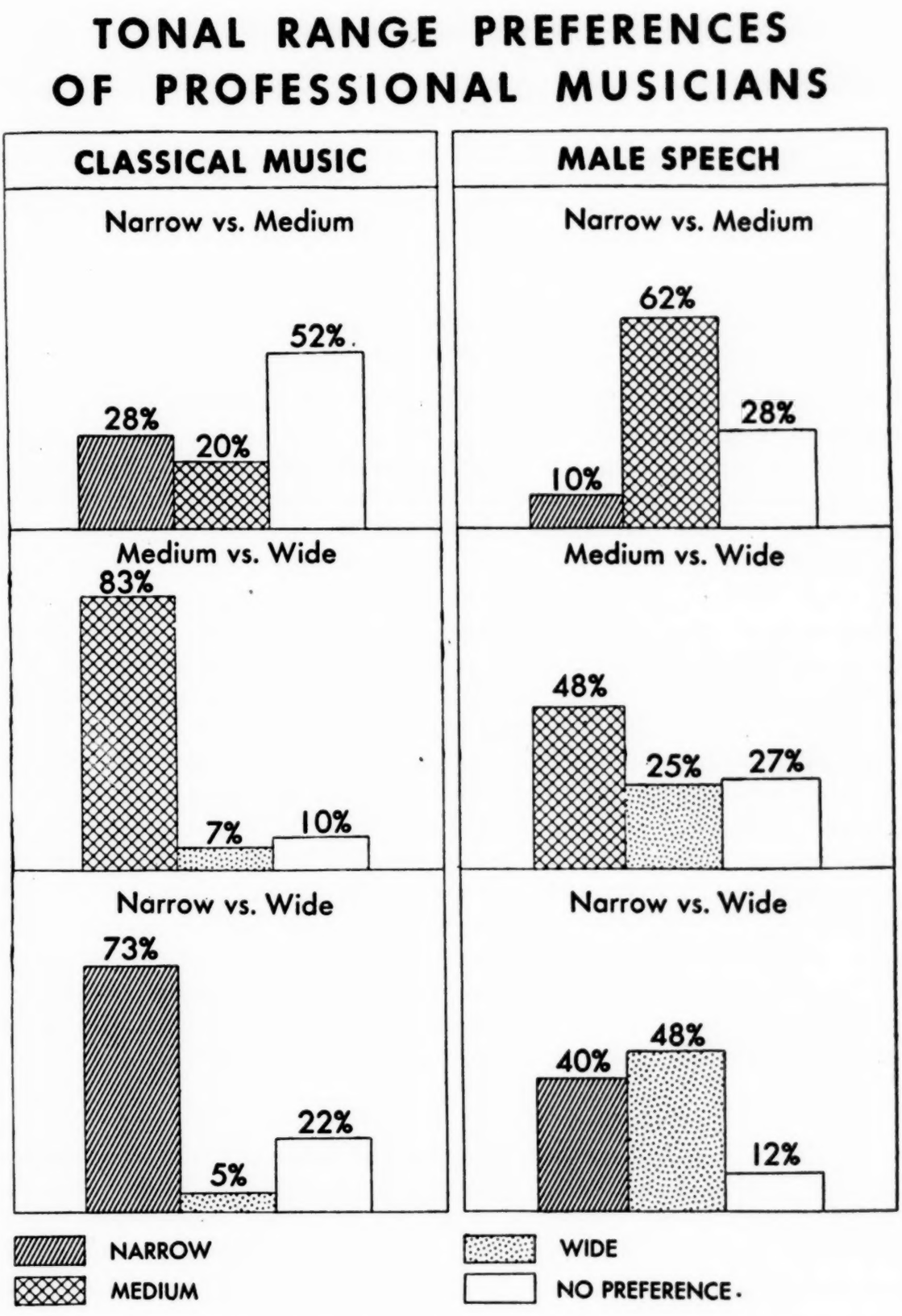



FIG. 3. The tonal range preferences of a group of professional musicians for classical music and male speech

TABLE V

TONAL RANGE PREFERENCES

*Listeners informed of conditions*

| Tonal Range | Experiment 1 (Cross-Section) | | Experiment 2 (Cross-Section) | | Experiment 3 (Musicians) | | Experiment 4 (FM Listeners) | |
|---|---|---|---|---|---|---|---|---|
| | Classical Music (B) | Male Speech (B) | Popular Music (D) | Mixed Speech (D) | Classical Music (B) | Male Speech (B) | Classical Music (B) | Male Speech (B) |
| Narrow | 55% | 55% | 34% | 44% | 80% | 63% | 31% | 48% |
| Wide | 32 | 34 | 37 | 34 | 15 | 32 | 55 | 46 |
| No Preference | 13 | 11 | 29 | 22 | 5 | 5 | 14 | 6 |
| Size of Sample | 105 persons | | 123 persons | | 20 musicians | | 96 persons | |

for the wide band indicates that some of the *S*s were impressed by the term 'high fidelity.'

The only group who reversed this trend were the FM listeners in *Experiment 4*. They showed a greater preference for 'high fidelity' to the 'low' than any other group when informed of conditions. In fact, the more they listen to FM, the more they indicated preference for high fidelity, but only *after* they had been acquainted with the conditions of the test. Since these special listeners did not show these preferences when they were unaware of the conditions of the tests, it seems that they were more influenced by the prestige of the term 'high fidelity' than any other group.

A point that lends further credence to this conclusion is that FM listeners, after being informed of the conditions, show an even more marked preference with music for the 'high fidelity' condition than with speech. This was the most marked reverse for preferences made as compared with choices *before* the *S*s were informed of the conditions. This finding seems to indicate that FM set owners have been susceptible to the statements that have been made stressing the high quality of *musical programs* on FM. The uninformed and informed preferences of FM listeners as compared with the cross section of listeners in Experiment 1 are illustrated in Fig. 4.

*Volume level preferences*

In *Experiment 6* there were four pairs of volume level comparisons; between low and moderate, between moderate and high volumes, first at a wide tonal range and then at a narrow tonal range (Table III, Pairs 6 to 9). As before, the results for both Music A and B and for both Speech A and B were combined. These results are presented in Table VI.

There was a marked preference for the moderate (60 db) volume level as against the low (50 db) level either at a narrow or a wide tonal range (Pairs 6 and 8). But there was no clear choice between

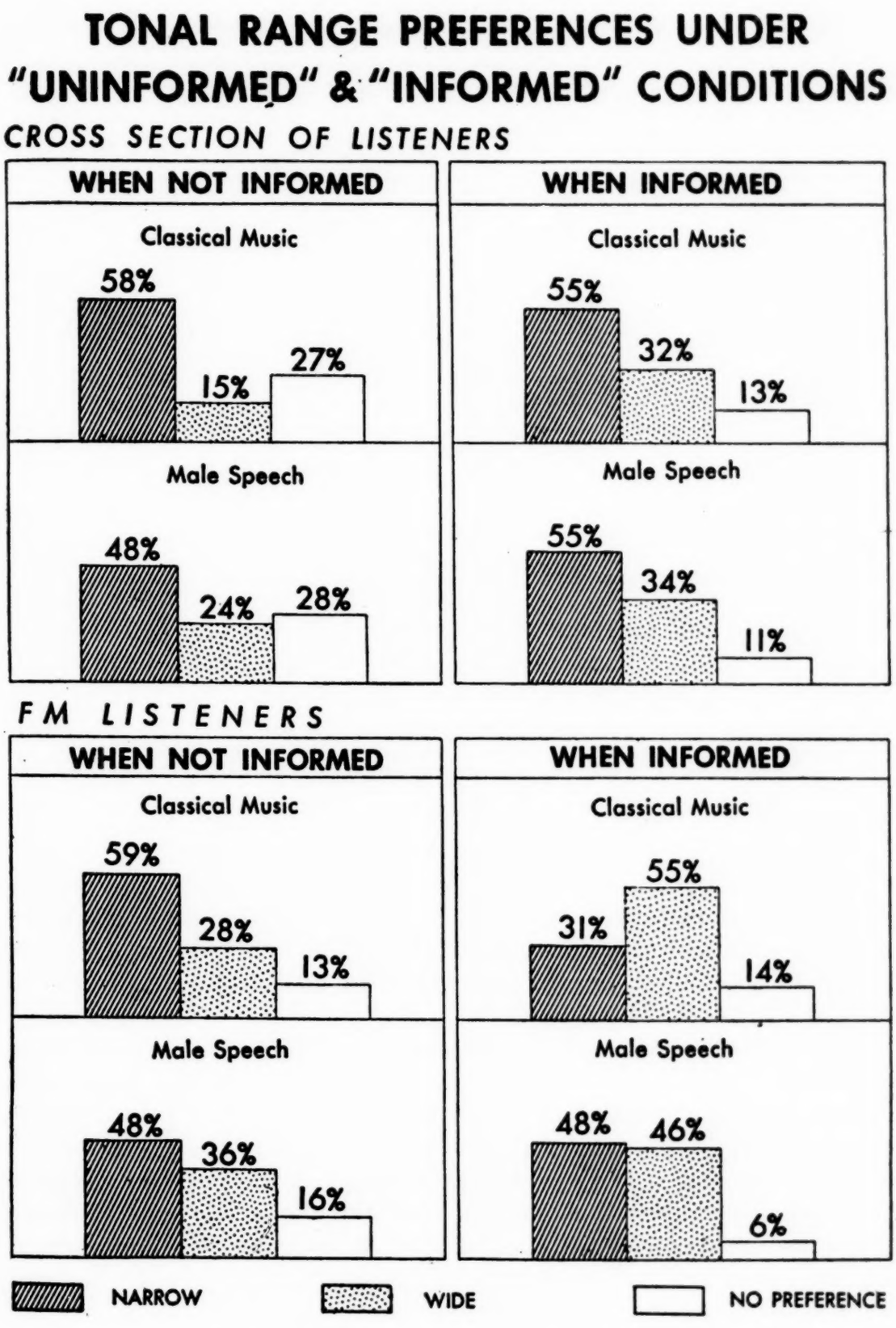



FIG. 4. The tonal range preferences of a cross-section of listeners and FM listeners for classical music and male speech when not informed and when informed that one part of a pair was narrow tonal range and the other wide tonal range

TABLE VI
VOLUME LEVEL PREFERENCES

| | Volume Level | Tonal Range | Experiment 6 | |
|---|---|---|---|---|
| | | | Classical Music (A & B) | Male Speech (A & B) |
| Pair 6 | Low | Wide | 9% | 2% |
| | Moderate | Wide | 72 | 79 |
| | No Preference | | 19 | 19 |
| Pair 7 | Moderate | Wide | 27% | 42% |
| | High | Wide | 46 | 38 |
| | No Preference | | 27 | 20 |
| Pair 8 | Low | Narrow | 18% | 4% |
| | Moderate | Narrow | 66 | 91 |
| | No Preference | | 16 | 5 |
| Pair 9 | Moderate | Narrow | 49% | 28% |
| | High | Narrow | 30 | 52 |
| | No Preference | | 21 | 20 |

(Size of sample: 111 persons)

the moderate and the high (70 db) levels (Pairs 7 and 9). In other words, the volume level preference for most listeners seems to lie somewhere between 60 and 70 db. In general, the volume level preference for speech seems to be somewhat louder than for music, particularly for narrow band reproductions (Pairs 8 and 9).

*Tonal range and volume level preferences*

To complete the study, experiments were undertaken to learn what the preferences would be when both tonal range and volume were varied simultaneously. As a part of *Experiments 1 through 5*, two such comparisons were undertaken (Table III, Pairs 4 and 5), and in *Experiment 6*, one (Table III, Pair 10). The results of this study are given in Table VII.

It is evident that in all three comparisons the narrow tonal range is preferred to the wide. However, the strength of the preferences depends upon the mixture of the elements. When in Pair 10, narrow tonal range at moderate volume (both of which are preferred elements) is compared with wide tonal range at low volume (both of which are non-preferred elements), then the choice is overwhelmingly for the preferred combination. In other cases, where the comparisons were not between definitely preferred and non-preferred elements, the choices were not as consistent or as clear cut.

*Group differences in preference*

The tonal range and volume preferences of the listeners of Experiments 1 and 6 were analyzed by sex, age, education, amount of

## TABLE VII

### Tonal Range and Volume Level Preferences

| | Tonal Range | Volume Level | Experiment 1 (Cross-Section) | | Experiment 2 (Cross-Section) | | | | Experiment 3 (Musicians) | | Experiment 4 (FM Listeners) | | Experiment 5 (Live Talent) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Classical Music (A & B) | Male Speech (A & B) | Piano Music (C) | Pop. Music (D) | Female Speech (C) | Mixed Speech (D) | Classical Music (A & B) | Male Speech (A & B) | Classical Music (A & B) | Male Speech (A & B) | Light Classical Music (I to K, M & N) |
| Pair 4 | Wide | Moderate | 21% | 18% | 42% | 41% | 35% | 24% | 3% | 5% | 13% | 19% | 22% |
| | Narrow | High | 62 | 70 | 51 | 38 | 52 | 61 | 67 | 75 | 79 | 72 | 73 |
| | No Preference | | 17 | 12 | 7 | 21 | 13 | 15 | 30 | 20 | 8 | 9 | 5 |
| Pair 5 | Wide | High | 28% | 43% | 55% | 18% | 41% | 46% | 10% | 30% | 40% | 45% | 47% |
| | Narrow | Moderate | 56 | 34 | 32 | 67 | 37 | 39 | 38 | 28 | 48 | 39 | 37 |
| | No Preference | | 16 | 43 | 13 | 15 | 22 | 15 | 52 | 42 | 12 | 16 | 16 |
| | Size of Sample | | 105 persons | | 123 persons | | | | 20 persons | | 96 persons | | 43 person |

| | Tonal Range | Volume Level | Experiment 6 (Cross-Section) | |
|---|---|---|---|---|
| | | | Classical Music (A & B) | Male Speech (A & B) |
| Pair 10 | Wide | Low | 10% | 2% |
| | Narrow | Moderate | 76 | 90 |
| | No Preference | | 14 | 8 |
| | Size of Sample | | 111 persons | |

musical training, musical preferences, and whether or not the subjects played an instrument. The criterion of a difference was determined not only by the size of the difference but also by its consistency.

In general, the analyses reveal no differences for any of these breakdowns, with one possible exception. Women display a consistent but *very slight* tendency to prefer wider tonal range than men. This trend is somewhat compensated for by a greater amount of 'no preferences' for women.

It is interesting that those over 26, when the average listener begins to suffer a hearing loss (**8**), show no difference in preference from younger listeners. Education plays no apparent role in preference; nor does musical training or experience. Furthermore, the results of the tests with the professional musicians and the FM listeners substantiate the finding that group differences, such as have been investigated, do not influence tonal range or volume level preferences. Such uniformity in results for various types of groups indicate no need for larger samples than were used.

*Reliability of judgments*

Three figures were found to be useful indices of reliability: (1) extent of strong preferences, (2) extent of no preferences, i.e., equal judgments, and (3) the degree of identity of choices on repeated pairs in which the content and the conditions of presentation were the same. Table VIII presents the results for music and speech by tonal comparisons, volume level comparisons, and by the combined type of comparisons for all six experiments.

It was found that the three indices were related to each other. Strong preferences and identity of choice were positively related, that is, the greater the strong preference, the greater the identity of choice. The extent of no preference was negatively related to the other two indices, that is, the greater the no preference, the fewer the strong preferences, and the smaller the identity of choice.

The tests in which both volume level and tonal range were varied together were the most reliable. The *S*s indicated strong preferences more often, no preference less often, and made the same judgment on repeated tests more often. This indicates in another way that both volume level and tonal range influence the judgment. When both factors vary, it is easier to make a reliable judgment than when only one factor varies. However, it should be noted that the volume judgments were more reliable than the tonal range judgments. This seems to indicate that, although both factors influence the judgment, listeners can distinguish more readily between different volume levels than between different tonal ranges.

In this connection, it should be noted that the tonal range preferences were not too reliable, even without comparison with other types of judgment. There was a relatively small amount of strong preference (a median for all five experiments of 9 percent); a relatively large amount of no preferences (median of 28 percent); and a relatively small amount of identical choices (median of 51 percent).

TABLE VIII

INDICES OF CONFIDENCE AND RELIABILITY

| Test | Experiment 1 (Cross-Section) | | Experiment 2 (Cross-Section) | | Experiment 3 (Musicians) | | Experiment 4 (FM Listeners) | | Experiment 5 (Live) | Experiment 6 (Cross-Section) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Music | Speech | Music | Speech | Music | Speech | Music | Speech | Music | Music | Speech |
| *Percentage of strong preferences* | | | | | | | | | | | |
| Volume Level | — | — | — | — | — | — | — | — | — | 12% | 22% |
| Tonal Range | 13% | 12% | 5% | 9% | 10% | 8% | 9% | 16% | 7% | — | — |
| Volume—Tonal Range | 22 | 17 | 20 | 11 | 14 | 19 | 32 | 27 | 17 | 19 | 27 |
| *Percentage of no preferences* | | | | | | | | | | | |
| Volume Level | — | — | — | — | — | — | — | — | — | 21% | 16% |
| Tonal Range | 30% | 25% | 44% | 31% | 28% | 23% | 26% | 18% | 35% | — | — |
| Volume—Tonal Range | 16 | 18 | 14 | 16 | 41 | 31 | 10 | 13 | 11 | 14 | 8 |
| *Percentage of identical choices* | | | | | | | | | | | |
| Volume Level | — | — | — | — | — | — | — | — | — | 48% | 63% |
| Tonal Range | 58% | 55% | 40% | 40% | 52% | 47% | 51% | 56% | 44% | — | — |
| Volume—Tonal Range | 59 | 64 | 47 | 63 | 63 | 63 | 70 | 71 | 58 | 65 | 80 |

This would seem to indicate that the listeners' judgments of tonal ranges were neither stable nor strong because they could not readily distinguish between different ranges; the situation was not as well structured as it was with volume choices or when both factors varied.

The judgments for Experiment 2 (variety of content) and Experiment 5 (broadcast music) were less reliable than in the other experiments. In the broadcast music experiment, this could be accounted for because of lack of control of content. However, in the variety of content experiment, where content was as well controlled as in the other experiments, another factor must be considered. It was found that as the preferences approach a medium tonal range, the listener does not make as clear-cut choices. One range sounds as pleasant to him as another. Therefore, he is more apt to shift his judgment on repeat tests.

## Summary

The main conclusions of this study are:

1. Listeners prefer either a narrow or medium tonal range to a wide one. However, the exact choice of bandwidth varies to some extent, within these limits, for different types of program content. Listeners prefer a slightly wider band for female speech, piano and popular orchestra than for male speech, mixed dramatic speech and classical orchestra.

2. Most listeners are not influenced to change their judgments when informed that one condition is 'low fidelity' and the other 'high fidelity'; under these conditions they still prefer a narrow to a wide tonal range. However, FM listeners were found to be more susceptible to this prestige suggestion than any other group.

3. Listeners prefer a peak sound intensity level somewhere between 60 and 70 db above the acoustical reference level. They prefer a higher volume level for male speech than for classical music.

4. Both tonal range and volume level influence the listeners' choices. The most preferred combination, within the limits of this study, seems to be a narrow tonal range and a 60 to 70 db acoustical level.

5. There were no clear-cut and consistent differences between various sub-groups of the audience: sex, age, education, amount of musical training, musical preferences, or whether they played a musical instrument.

6. Two groups with special experience (professional musicians and FM listeners) indicated substantially the same preferences as did the cross section of listeners.

7. The most reliable judgments were made when both tonal range and volume level were varied. However, judgments for volume alone were more reliable than for tonal range alone. Judgments were least reliable when content was not controlled and when preferences were for a medium tonal range band.


## Discussion

Contrary to the expectation of most engineers, it was found that listeners prefer either a narrow or medium tonal range band to a wide band when listening to reproduced sound. It may be argued however, that this preference for restricted tonal range is the result of years of experience of listening to a narrow band on the radio and on recordings. The argument generally continues that with experience, listeners will overcome this initial dislike and discover new enjoyment in wide range. It is important to understand that this argument is based on no published evidence whatsoever. It is an hypothesis, unfortunately often stated as a fact. Actually, listeners have had

extensive experience with wide range in speech, because that is what they hear all day long in ordinary conversation. Yet, despite their experience, they did not prefer a wide band for speech in these tests. In addition, listeners have experienced a relatively wide range in the sound produced in motion picture theatres.

With regard to music, it has been found that listeners prefer narrow range no matter how much musical training and experience they have had. Even professional musicians, who have had long years of experience with music produced at the widest tonal range, still manifest the same preferences as other listeners.

It is admitted that these experiments do not furnish conclusive evidence to settle this question one way or the other. The data however, tend to confirm the alternative hypothesis: that a narrow tonal range is preferred because it sounds better, and not because the listeners' tastes have been spoiled. In other words, a 'natural' pleasantness theory is favored to a conditioned response theory.

The factor of suggestion is also an important consideration in this connection. It was found in these experiments, that though listeners do not prefer a wide tonal range, the entire matter of range is probably not as important to them as is volume level. When volume level was controlled, the choices were not made with great confidence as shown by the relatively small amount of strong preferences, by the high percentage of no preference judgments, and by the frequent changes of judgment on repeat tests. Apparently, then, listeners cannot discriminate as well between differences in tonal range as they can with volume level. Certainly tonal range is not as important as freedom from background noise. Because tonal range is generally not a matter of great concern to most listeners, and because strong preferences are not evident, it was possible to influence some listeners to choose a wide range when they were told the presentation was 'high fidelity.' In other words, it was possible to influence listeners when the situation was unstructured. On the other hand, it would be much more difficult to influence listeners in a structured situation. For example, it is unlikely that verbal suggestions would influence listeners to choose volume levels they disliked or presentations with excessive background noise.

If, at some future time, there should be a predominant shift to a preference for wide range, the question may well be raised as to the cause of the change. Would it be the result of greater listening pleasure through wide range experience, or would it be simply the prestige suggestion effect resulting from the use of the terms 'wide range' and 'high fidelity'?

(Manuscript received February 24, 1945)

## Bibliography

1. Chinn, H. A. Broadcast studio audio-frequency systems design. *Proc. Inst. Radio Eng.*, 1939, **27**, 83–87.
2. Fletcher, H. Auditory patterns. *Rev. mod. Physics*, 1940, **12**, 47–65.
3. Fletcher, H. Hearing, the determining factor for high fidelity transmission. *Proc. Inst. Radio Eng.*, 1942, **30**, 266–277.
4. Gannett, D. K., & Kerney, I. Discernability of changes in program bandwidth. *Bell Sys. tech. J.*, 1944, **23**, 1–10.
5. Hanson, O. B. Comments on high fidelity. *Electronics*, 1944, **17**, No. 8, 130–131, 385–391.
6. Smeby, L. C. Recording and reproducing standards. *Proc. Inst. radio Eng.*, 1942, **30**, 355–356.
7. Snow, W. B. Audible frequency ranges of music, speech and noise. *J. acoust. Soc. Amer.*, 1931, **3**, 155–166.
8. Steinberg, J. C., Montgomery, H. C., & Gardner, M. B. Results of the World's Fair hearing tests. *Bell Sys. tech. J.*, 1940, **19**, 533–562.

# A SENSITIZED EYELID REACTION RELATED TO THE CONDITIONED EYELID RESPONSE[1]

BY DAVID A. GRANT

*University of Wisconsin*

The present study was designed to investigate factors related to the sensitization of secondary eyelid responses to light. Secondary responses were defined as those having latencies ranging from 120 to 240 msec. The secondary responses would thus occur after the primary reflex to light (latency 40–110 msec.) when the latter occurs. Sensitization of the response is shown by reliable increases in frequency and amplitude.

Previous studies (**10, 11, 14, 13, 1**) have shown the eyelid reaction to light to be complicated, but no consistent secondary response has been reported.

In a previous experiment aimed at investigating pseudo-conditioning of the eyelid response (**6**) the present writer obtained what appeared to be pseudo-conditioned responses to light. These responses had latencies greater than those of eyelid reflexes to light. The behavior of similar secondary eyelid responses in a control group forced the writer to conclude that sensitization processes other than pseudo-conditioning were involved. This conclusion was supported by the findings of a second experiment (**7**). In both studies the frequency of eyelid closures to light was observed to increase markedly over a period during which no other stimulation was given. The increments did not, apparently, involve either conditioning or pseudo-conditioning, although in latency and other respects the responses involved resembled conditioned eyelid responses.

The latency distributions of the secondary eyelid responses were so strikingly similar in the two earlier studies that it was apparent that some sort of consistent response organization was involved and that further research was desirable to determine its properties. The probable significance for conditioning and reinforcement theory of such a response organization made the need for further analytic study more urgent. The present research was designed to investigate some of the characteristics of this secondary response organization.

[1] This research was supported in part by a grant-in-aid from the Graduate Research Fund of the University of Wisconsin for the year 1943–44.

## Apparatus and Procedure

The apparatus used in the present study was essentially the same as that used in a previous study of eyelid responses (8). It consisted of a modification of the Dodge photochronograph (2) arranged to record movements of the right eyelid by means of the shadow of a simple lever system.

The only stimuli used were light stimuli which consisted of increases in the brightness of a circular milk-glass stimulus screen, 10 cm. in diameter, located 60 cm. directly in front of the *S*'s eyes. In order to determine the influence of stimulus intensity on the secondary responses, electric bulbs of four different wattages were used. These gave brightness increases to 4.9, 46, 182, and 582 millilamberts without marked shifts in hue and are hereafter referred to as $L_1$, $L_2$, $L_3$, and $L_4$, respectively. (All photometry was carried out by means of a Macbeth Illuminometer.) The duration of all stimuli was about 750 milliseconds.

Experiments were conducted in the afternoon in a sound-proofed, air-conditioned laboratory. After each *S* was adjusted to the apparatus, instructions were read to him which indicated that he was to view binocularly a fixation spot in the center of the milk-glass stimulus screen, and that he was not to attempt to control or alter his natural responses to the stimuli. He was then run through the following procedure:

*Test-Series 1* ($L_1$, $L_2$, $L_3$, $L_4$, and one control trial)
Fixation (*S* looks at screen for 8 min.)
*Test-Series 2* (Like Test-Series 1, above)
Rest (2 min.)
Fixation (Like Fixation, above)
*Test-Series 3* (Like Test-Series 1, above)
Rest (2 min.)
Fixation (Like Fixation, above)
*Test-Series 4* (Like Test-Series 1, above)

In each of the above test series, the four light stimuli were given in permuted sequences which varied from series to series and from *S* to *S* in a pre-determined randomized order. The control trials during each test series consisted of recording eyelid movements while all of the apparatus was operated, except that no stimuli were given. The interval between trials within a series varied randomly from 25 to 45 sec., and no ready signals were given. The use of four test series, spaced throughout the experiment, permitted investigation of the rate of sensitization.

The *S*s were men and women students who volunteered from elementary courses in psychology during the summer of 1944. The *S*s were assigned to three groups according to their order of appearing for the experiment. There were 20 *S*s in each group, and the records of an additional four *S*s were disqualified because they included three or more random blinks during control trials. The procedure used with the three groups varied only in the degree of pre-stimulus illumination. For Group A, the pre-stimulus illumination was about .01 millilamberts, for Group B, it was about .10 millilamberts, and for Group C, it was about .55 millilamberts. The use of the three groups permitted determination of the influence of pre-stimulus illumination on the sensitization of the eyelid response.

## Results

1. *Latency of secondary eyelid responses.*—Single or multiple eyelid responses were evoked by the light stimuli from most of the *S*s. To aid in the analysis all obtained responses were plotted, frequency against latency,[2] in Fig. 1. This distribution is distinctly bimodal, and it agrees in all but inconsequential details with the corresponding distributions reported earlier (6, 7).

[2] Latencies in the present study have been corrected for the latency of recording equipment and may be considered absolute. This correction was not made in earlier studies (6, 7), so that for purposes of comparison it is necessary to subtract 50 msec. from the recorded latencies given in the earlier reports.

The portion of the latency distribution ranging from 40 to 110 msec. is composed largely of primary reflexes to light (9). These primary reflexes will be referred to in this paper as *α-responses.* The portion of the distribution ranging from 120 to 240 msec., with a mode at 140–150 msec., contains the secondary responses which are

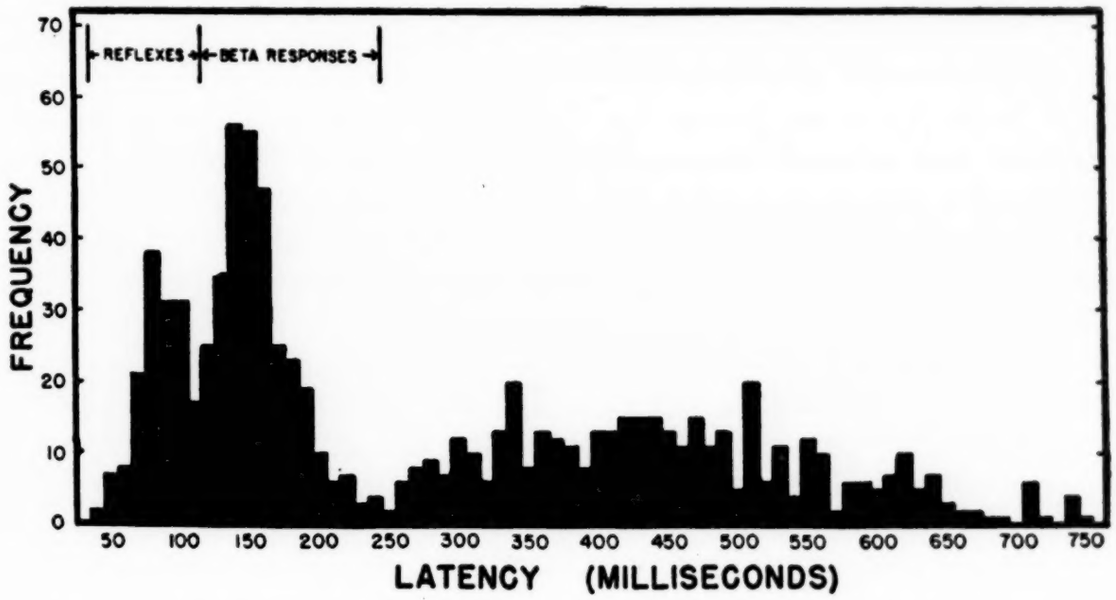



FIG. 1. Distribution of the latencies of all eyelid responses evoked by light stimuli during the various test series of the experiment

the chief concern of this experiment. They will be referred to as *β-responses.* Naming these responses is justified by two considerations: (*a*) the β-response exhibits consistent latency characteristics under differing experimental conditions as is shown by comparing present and previously reported latency distributions (see above); and (*b*) the β-response differs functionally from the α-response (9) and from random blinking (6) in that it undergoes reliable *increment* in frequency during periods of visual fixation while no obvious conditioning reinforcement was being given (6, 7).

TABLE I

FREQUENCY OF BETA-RESPONSES DURING THE DIFFERENT TEST-SERIES OF THE STUDY

| | Group A | | | | | Group B | | | | | Group C | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test-Series | *1* | *2* | *3* | *4* | *Tot.* | *1* | *2* | *3* | *4* | *Tot.* | *1* | *2* | *3* | *4* | *Tot.* |
| Stimulus $L_1$ | 1 | 4 | 3 | 2 | 10 | 2 | 3 | 3 | 4 | 12 | 0 | 0 | 3 | 3 | 6 |
| Stimulus $L_2$ | 2 | 6 | 10 | 10 | 28 | 1 | 7 | 8 | 10 | 26 | 0 | 4 | 4 | 7 | 15 |
| Stimulus $L_3$ | 5 | 10 | 7 | 11 | 33 | 4 | 10 | 12 | 12 | 38 | 2 | 8 | 11 | 10 | 31 |
| Stimulus $L_4$ | 5 | 12 | 13 | 12 | 42 | 5 | 6 | 8 | 13 | 32 | 9 | 11 | 11 | 10 | 41 |
| Total | 13 | 32 | 33 | 35 | 113 | 12 | 26 | 31 | 39 | 108 | 11 | 23 | 29 | 30 | 93 |

The upper latency limit (240 msec.) of β-responses seemed reasonably definite in Fig. 1. The lower limit (120 msec.) was, however, less clear-cut, presumably because there was some overlapping of the distribution of reflex α-responses with the distribution of β-responses at that point. Use of these limits will, however, probably introduce no important errors.

2. *Frequency of β-responses as a function of the intensity of stimulus and pre-stimulus illumination.*—The number of β-responses evoked from Ss of the three groups by the four light stimuli during each of the four test series of the experiment is presented in Table I.

Comparison of the total frequency for each test series showed a general increment during the course of the experiment. The totals also showed that the number of responses was influenced by stimulus intensity.

The frequency data of Table I are presented graphically in Figs. 2, 3, and 4 to facilitate analysis. For interpretive convenience both frequency and percent frequency scales are given. In Fig. 2 the frequency of β-responses is plotted for each group against the logarithm of stimulus intensity. Frequency is seen to be influenced decisively by stimulus intensity. Statistical reliability of trends will be discussed in the next section.

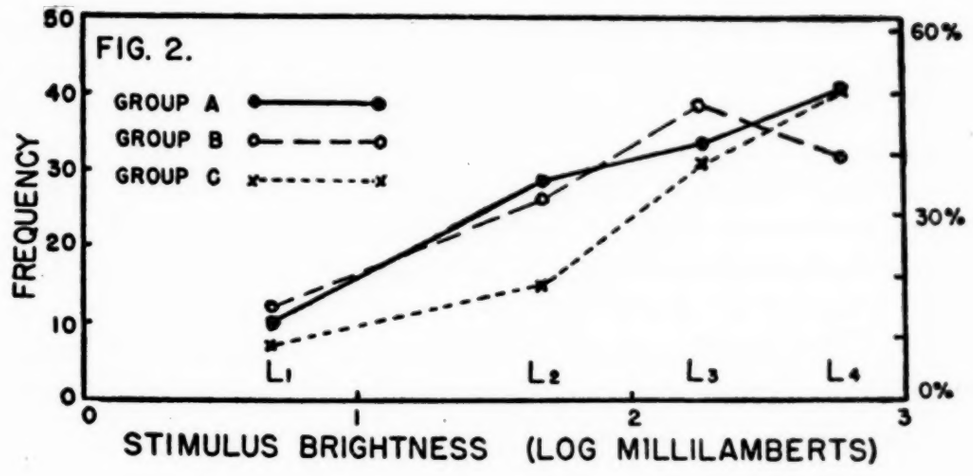
In Fig. 3 the frequency of β-responses is plotted for each stimulus against the logarithm of the intensity of the pre-stimulus illumination. Although there is a slight general downward trend for frequency as pre-stimulus illumination increases, the trend is neither sharp nor consistent.

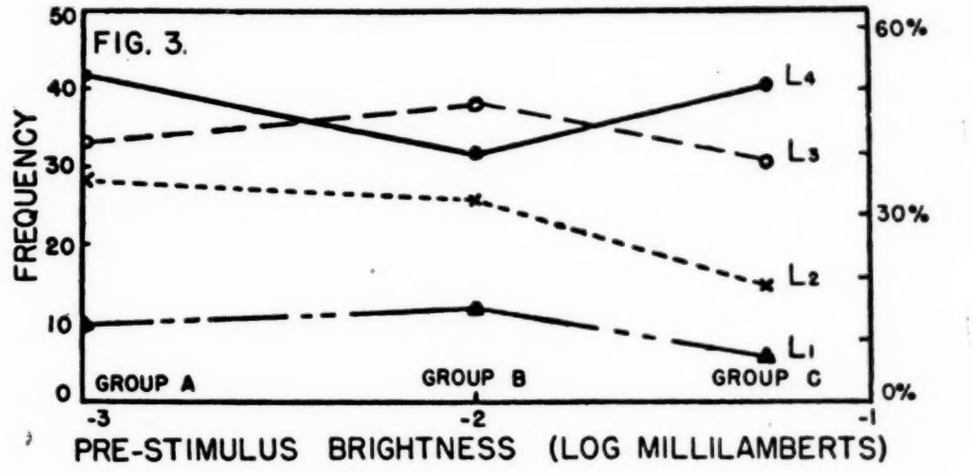
TABLE II

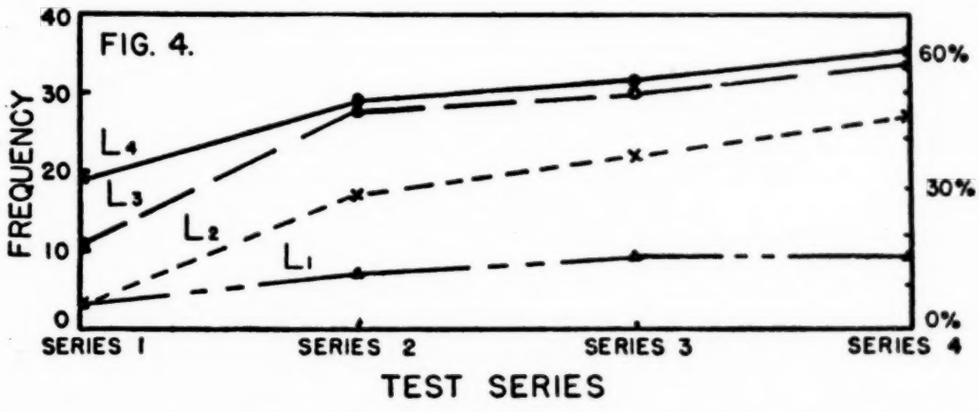
ANALYSIS OF VARIANCE OF THE FREQUENCY OF BETA-RESPONSES

| Source | Sum of Squares | *df* | Mean Square | Variance Ratio | *P* |
|---|---|---|---|---|---|
| Stimulus | 377.0834 | 3 | 125.6945 | 40.7358 | < .1% |
| Series | 222.7500 | 3 | 74.2500 | 24.0634 | < .1% |
| Pre-Stimulus | 13.5417 | 2 | 6.7708 | 2.1943 | <25% |
| Stim.-Series | 38.0833 | 9 | 4.2315 | 1.3714 | <50% |
| Stim.-Pre-Stimulus | 37.2916 | 6 | 6.2153 | 2.0143 | <20% |
| Series-Pre-Stimulus | 9.6250 | 6 | 1.6042 | 1/1.9234 | |
| Stim.-Series-Pre-Stim. | 55.5417 | 19 | 3.0856 | | |
| Total | 753.9167 | 47 | | | |

3. *Sensitization of β-responses.*—In Fig. 4 the frequency of β-responses is plotted for each stimulus against test series, thus showing the increasing frequency (response sensitization) which occurred during the experiment. In Fig. 4 it is apparent that as in previous studies (**6, 7**) the frequency of responses to each stimulus increased




FIGS. 2, 3, and 4. Relationships between frequency of $\beta$-responses and stimulus brightness (Fig. 2), pre-stimulus brightness (Fig. 3), and test series (Fig. 4)

markedly during the experiment without obvious conditioning or pseudo-conditioning reinforcement. The frequency of response to $L_1$ increases less than does the frequency of response to $L_2$, $L_3$, $L_4$.

In order to test the statistical significance of the trends described in this and the preceding section, an analysis of variance was performed on the data of Table I. The analysis is summarized in Table II in the conventional manner (3), and the percent confidence levels,

as determined from Fisher and Yates (4) and Merrington and Thompson (15), are indicated under *P* for each variance ratio. From Table II it is obvious that the influence of stimulus intensity (Stimulus) on frequency of β-responses was highly reliable. The sensitization trend (Series) was also highly reliable. The influence of pre-stimulus illumination (Pre-Stimulus) on response frequency was not statistically significant, nor were any of the second-order interactions.

4. *Degree of sensitization of β-responses as a function of the intensity of stimulus and pre-stimulus illumination.*—The reliable frequency increments shown in Fig. 4 constituted definite proof of the sensitization of the β-responses under the conditions of the present investigation. A good index of sensitization would be the overall gain in frequency from Test Series 1 to Test Series 4. Such indices for each group and each stimulus intensity could be obtained by subtracting the entries in Table I under Test Series 1 from the corresponding entries under Test Series 4. This has been done to supply the entries for Table III. Table III shows that the frequency

TABLE III

INCREASES IN FREQUENCY OF BETA-RESPONSES FROM TEST SERIES 1 TO TEST SERIES 4

| | Stimulus Intensity | | | | Total |
|---|---|---|---|---|---|
| | $L_1$ | $L_2$ | $L_3$ | $L_4$ | |
| Group A | 1 | 8 | 6 | 7 | 22 |
| Group B | 2 | 9 | 8 | 8 | 27 |
| Group C | 3 | 7 | 8 | 1 | 19 |
| Total | 6 | 24 | 22 | 16 | 68 |

increments differed according to the stimulus intensity—at least the gains for $L_1$ were consistently less than the gains for $L_2$ and $L_3$.

An analysis of the variance of the data in Table III revealed that, in general, the influence of stimulus intensity on gains in frequency was significant at the 5 percent confidence level. Variation from Group A to Group C in pre-stimulus illumination produced no significant over-all differences in gains. Further analysis of the data of Table III by use of Student's *t*-test (3) showed that the difference in gains between $L_1$ and $L_2$ was significant at the 1 percent confidence level, and the difference between $L_1$ and $L_3$ was significant at the 3 percent confidence level. The degree of sensitization of β-responses is thus shown to be a function of stimulus intensity.

5. *Amplitude of β-responses.*—An analysis of the amplitude of the β-responses was made in the same general way that the frequency

data were analyzed. Although the amplitude data were more irregular than the frequency data, the mean amplitude increased reliably (1 percent confidence level) during the course of the experiment, but the amplitude of $\beta$-responses was not significantly influenced by different intensities of stimulus or pre-stimulus illumination. Recorded amplitude increased from an average of about 3.7 mm. on Test Series 1 to about 9.6 mm. on Test Series 4. (The lid movements were magnified threefold by the lever system.) The significant increase in average amplitude is further indication of the sensitization of the $\beta$-responses.

6. *Frequency and amplitude of the $\alpha$-response (primary reflex to light).*—The frequency of primary reflexes was low in the present experiment, and—possibly because of the small number of reflexes—frequency could not be demonstrated to be affected by stimulus or pre-stimulus intensity nor were there reliable changes in frequency from test series to test series. The amplitude of the reflexes was probably affected in an overall manner by some of the experimental variables, as shown by an analysis of variance. The amplitude measures, however, revealed no consistent trends, and individual *t*-tests yielded no significant (5 percent level) differences.

The increases in frequency and amplitude which were characteristic of the $\beta$-responses were not duplicated with the $\alpha$-response. The two responses appeared, therefore, to be functionally discrete.

## Discussion

1. *The $\beta$-response of the eyelid to light.*—Discrete latency range and functional differences make it necessary to distinguish the secondary or *$\beta$-response* of the eyelid studied in the present investigation from the primary reflex or *$\alpha$-response* to light. The distinct latency characteristics which have appeared consistently in two other studies have been pointed out above in Section 1 under *Results*. The present study further verified that the frequency and amplitude increases of the $\beta$-response are not related to the frequency and amplitude trends of the $\alpha$-response. This finding harmonizes with unpublished findings of earlier studies (**6**, **7**). In one of the previous experiments (**5**), frequency, amplitude, and latency of secondary eyelid responses[3] were correlated + .04, + .25, and − .31, respectively, with the corresponding measures of the primary reflexes to light. (These correlations were based upon 31, 25, and 25 cases, respectively, and do not differ significantly from zero.) These differences have necessitated distinguishing between the $\beta$-responses and the primary reflexes or $\alpha$-responses.

[3] There termed 'Fixation Responses.'

2. *Significance of the β-response.*—The results of the present study have obvious implications for eyelid pseudo-conditioning. The role of the unconditioned stimulus (air-puff) in eyelid pseudo-conditioning was not clarified by previous studies (**6**, p. 148; **7**, p. 211). The present findings indicate clearly that the air-puff was not the sole cause of the frequency increments of those experiments and fully justify the suggestion (**6**, p. 146) that the apparent pseudo-conditioning results be ". . . referred to by means of some such general term as *sensitization* rather than as *pseudo-conditioning.*" The air-puff may have contributed to the sensitization (**7**, p. 211), but further experimentation is needed to determine more precisely the nature of the contribution.

The present findings have broader implications for eyelid conditioning and learning theory. The eyelid response has been the basis of some of the most precise studies of conditioning and learning (**12**), and findings of the present experiment should lead to more complete understanding of the reinforcement process in eyelid conditioning. In a previous study (**7**) it was stated that "the results give the impression that the observed differences (between conditioned, pseudo-conditioned, and secondary eyelid responses) were superimposed upon an underlying identity in response." Specifically, it might now be suggested that the conditioned eyelid response is originally a sensitized β-response which, during reinforcement trials, changes to fit the discriminated features of the experimental situation. Further research would be needed to prove or disprove this hypothesis.

One specific implication of the present findings arises from the fact that responses to stimuli of low intensity are not markedly altered. This suggests that light stimuli of low intensity be used to restrict sensitization effects when associative aspects of eyelid conditioning are to be emphasized. This may prove very helpful in future research directed at unravelling the effects of sensitization from those of association in eyelid conditioning.

3. *The nature of the β-response.*—The present findings raise the problem of the nature of the β-response and its sensitizing mechanism. The three most inclusive possibilities would be that the β-response is: (*a*) a voluntary response; (*b*) a preëxisting learned or conditioned response; or (*c*) a secondary reflex to light. The short latency plus the fact that *S*s are unable to report any self-instructed set (**5**) would seem to rule out the possibility of the β-response being an ordinary voluntary response. The behavioral facts do not fit the classical conception of either of the other alternatives, and indication of a choice is beyond the scope of the present paper. The present findings do, however, point up the likelihood that in the eyelid conditioning reinforcement is at least a *dual* process, involving non-associative

sensitization of a preëxisting response organization as well as the consequences of the association of the conditioned and unconditioned stimuli and their responses. Further knowledge of the nature of the effector system should inevitably result in more precise formulations of reinforcement theory and more adequate understanding of the learning process.

## Summary

The purpose of the present experiment was to investigate factors related to the sensitization of $\beta$-responses of the eyelid (secondary eyelid responses ranging in latency from 120–240 msec.) to light. Four degrees of stimulus intensity and three degrees of pre-stimulus intensity were varied systematically in three groups of 20 Ss each. Eyelid responses were recorded during four test series, spaced at intervals during a period of visual fixation. Stimuli were presented and eyelid movements were recorded by means of a modified Dodge photochronograph.

The findings of the experiment may be summarized as follows:

1. The $\beta$-response was found to have the same general latency distribution as reported in earlier studies.

2. The frequency of $\beta$-responses was found to be reliably affected by stimulus intensity but not by pre-stimulus intensity.

3. Reliable increases in frequency and amplitude of the $\beta$-response occurred during a period of visual fixation in the absence of any obvious conditioning or pseudo-conditioning reinforcement.

4. The degree of sensitization differed with stimuli of different intensity, being least striking with low illumination (5 millilamberts).

5. The changes in frequency and amplitude of $\alpha$-responses (primary reflexes to light) were not systematically related to the trend of the $\beta$-responses.

The findings of the present experiment support previously reported suggestions that apparent pseudo-conditioning of the eyelid response was some form of sensitization dependent on other than non-associative effects of the unconditioned stimulus.

It is conjectured that the conditioned eyelid response consists of the sensitized $\beta$-response altered according to the discriminated aspects of the stimulus pattern.

(Manuscript received March 12, 1945)

## References

1. Bender, M. B. Eyelid closure reaction. *Arch. Ophthal., Chicago*, 1943, **29**, 435–440.
2. Dodge, R. A pendulum photochronograph. *J. exp. Psychol.*, 1926, **9**, 155–161.
3. Fisher, R. A. *Statistical methods for research workers.* London: Oliver and Boyd, 7th edition, 1941.

4. FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver and Boyd, 1938.
5. GRANT, D. A. *Sensitization and association in eyelid conditioning.* Unpublished doctoral dissertation, Stanford Univ., 1941.
6. GRANT, D. A. The pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1943, **32**, 139–149.
7. GRANT, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.
8. GRANT, D. A., & ADAMS, J. K. 'Alpha' conditioning in the eyelid. *J. exp. Psychol.*, 1944, **34**, 136–142.
9. HILGARD, E. R. Conditioned eyelid reactions to a light stimulus based on the reflex wink to sound. *Psychol. Monogr.*, 1931, **41**, No. 184.
10. HILGARD, E. R. Reinforcement and inhibition of eyelid reflexes. *J. gen. Psychol.*, 1933, **8**, 85–113.
11. HILGARD, E. R. Modification of reflexes and conditioned reactions. *J. gen. Psychol.*, 1933, **9**, 210–215.
12. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton Century, 1940.
13. KÖDDING, I. Der Lidschlag im Kindesalter. *Mschr. Kinderheilk.*, 1940, **84**, 212–223.
14. LUCKIESH, M., & MOSS, F. K. Frequency of blinking as a clinical criterion of ease of seeing. *Amer. J. Ophthal.*, 1939, **22**, 616–622.
15. MERRINGTON, M., & THOMPSON, C. M. Tables of percentage points of the inverted Beta (F) distribution. *Biometrika*, 1943, **33**, 73–88.

# EVIDENCE FOR A UNIVERSAL SHOCK FACTOR IN LEARNING[1]

BY GEORGE S. SNODDY

In 1935 the author (4) presented experimental evidence for two opposed growth factors in the learning of the mirror vision coördination. In this type of learning, which can be brought under the most adequate experimental control, there appeared no doubt of the accuracy of our assumption regarding the two opposed growth processes, but it was not contended that this was necessarily true of all types of learning. In 1938 Doré and Hilgard (2) attacked the writer's work by assuming that if they could get results from another type of learning not in accord with the author's findings, the writer would be shown to be at fault. They secured results partly supporting the author's contention and partly their own assumption. Of course, the logic of their assumption is questionable since, after all, the findings of the author might have been due to the success of his experimental controls which their experiment had not duplicated; this, in fact, turns out to be the case. It would have been far better to repeat the writer's experimental setting, but this did not seem promising or practical since the writer's research involved years of work and close to a thousand learning *S*s. Bell (1), a student of Hilgard, repeated Hilgard's experiment and found further support for the author, but did not depart from the customary learning conceptions which the author's findings appear to contradict. Again there was no attempt to repeat the writer's experiment; and in Bell's experiment appear the same fatal defects in experimental control which are found in Hilgard's work.

Following the Hilgard attack, the writer in 1939 set about to alter his experimental approach in such manner as to recheck his previous findings and at the same time to determine the defects latent in Hilgard's learning controls. The experiments were completed in 1940 but due to the writer's ill health they have remained unpublished to the present time. The results seem to indicate very clearly that the two growth processes not only obtain in the mirror vision coördination, but apparently they must be universal in the entire field of learning, or wherever perception is involved.

[1] This research was completed at Indiana University in 1940. The writer is greatly indebted to Dean Fernandus Payne for appropriations for the research, and to Dr. W. W. Wright of the School of Education for providing the entire *B* group in our first experiment.

## The Mirror Vision Experiment

This experiment employs a star pattern cut from brass (Fig. 1) and mounted on a glass plate so that a path in the shape of a star is provided for tracing by a metal stylus in electric circuit with the brass plates. The path is shielded from direct vision and is traced by mirror reflection. The object is for the *S* to trace the path and avoid contact with the brass plates. Contacts are recorded by an electric counter. One of the unique advantages residing in this experiment is the excellent control of the *S*'s attitude. The errors, or contacts with the plates, are kept sta-

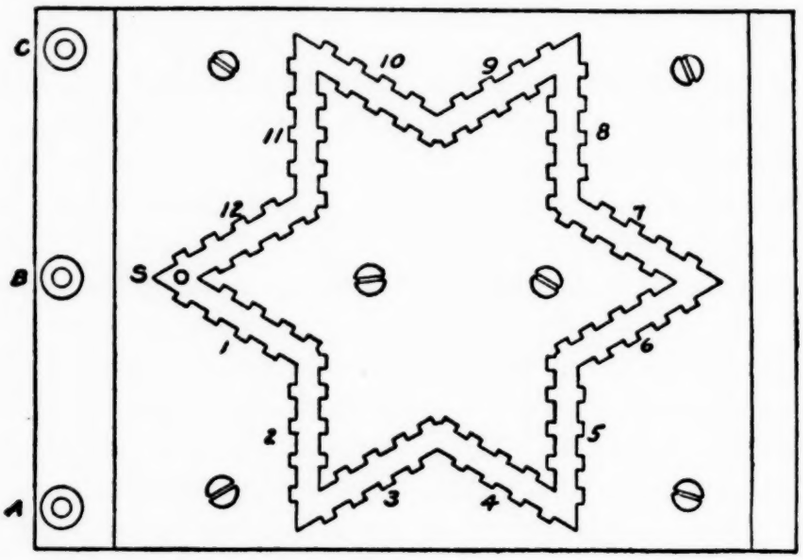



Fig. 1. The tracing instrument. A mirror is placed at the far side, and a screen is placed between the *S* and his direct vision of the instrument. The path is 500 mm. in length. The sides of the path are niched to prevent sliding of the stylus along the edges, thus learning of the coordination is quickly enforced.

tistically equal with the time in sec. Thus, an *S* laboriously struggling to avoid all errors would make too low a score due to the increase of the time consumed. The length of the path was 500 mm.; thus the score $\frac{500}{\frac{T+E}{2}} = \frac{1000}{T+E}$ provides an exact measure of the efficiency of the *S*, which reduces to millimeters of space covered per sec., with one error per sec. A table of reciprocals can be readily employed to turn these time and error scores into velocity per sec., with the error factor eliminated through reduction to unity.

## The Unusual Character of Mirror Coördination Findings

The results from our 1935 (4) experiments were striking, and many readers were of the opinion that they could be duplicated in other acts of learning, but this now appears to be somewhat in error. The outstanding feature of the mirror vision coördination was concerned with the demonstrated positive effect of the time intervals between repetitions. These were constructive to the highest degree, even after a minimum of practice. The prevailing logic which learning theory has evolved over the years, i.e., that the only constructive processes following activity are confined to the elimination of fatigue,

or as Hilgard says 'work effects,' is completely contradicted. One repetition of the apparatus per day gave an enormously superior efficiency *early in the learning;* thus, the time interval was not eliminating fatigue. It seemed that two growth processes were clearly involved: first, an early growth that was directly dependent upon the time intervals, and which varied directly with the length of the interval and second, another later growth which depended upon repetition without time interval. The first was permanent, *since*

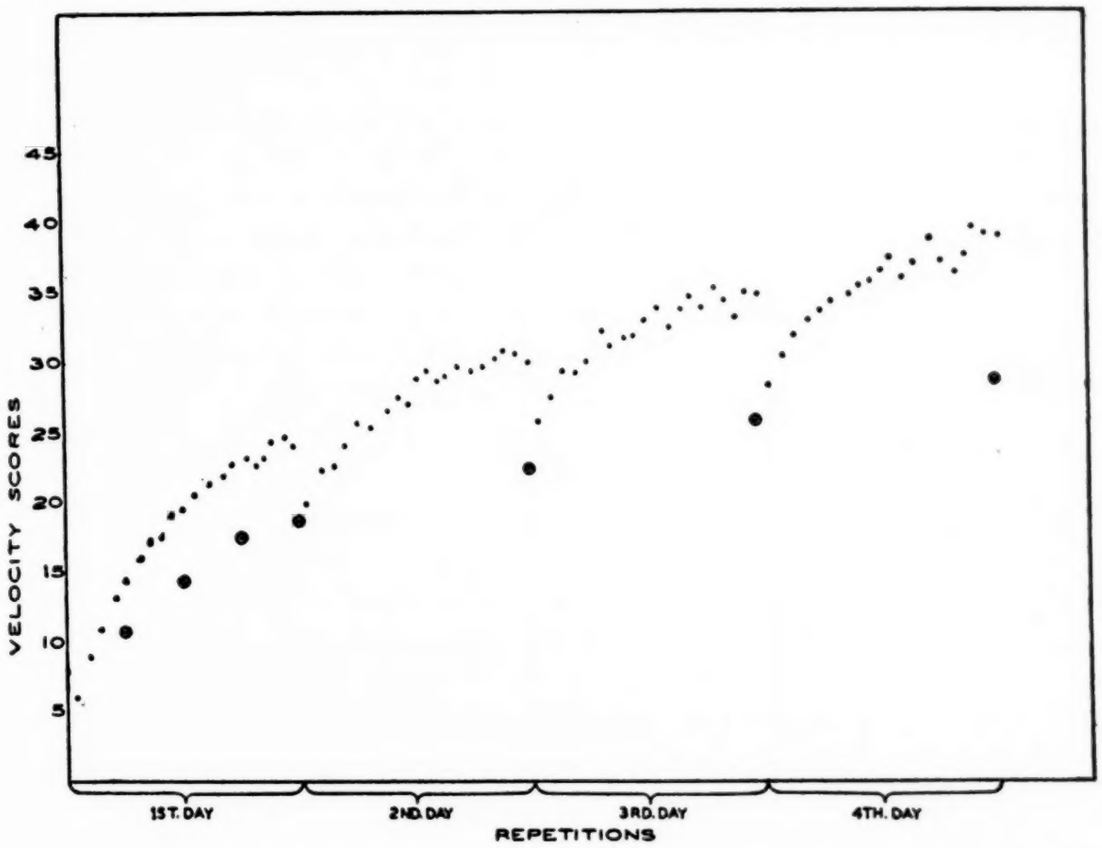



FIG. 2. Showing four days of practice with 20 circuits each day and one min. between circuits. The encircled stations are the cumulative means. The forgetting factor is shown by the loss between days of practice. This loss increases as practice continues, showing that the secondary or temporary growth is greater in the latter part of practice. Secondary growth is regarded as physiological summation.

*time would always increase it*, and the second was temporary and interpolated time always reduced it. Since early interpolations of time encourage a very large growth or increment, and later closely put repetitions a much smaller increment, it was apparent that the two growths are in large measure consecutive. If a practice consists of one repetition per day, that is with time intervals of 24 hours, a very large primary growth takes place; but this has no secondary or temporary growth at all, regardless of the length of practice over the days and weeks. The time intervals in this case determine but one

growth, the primary. But if the practice consists of 20 repetitions per day, Fig. 2, with one-min. intervals between repetitions, a good growth takes place during the practice periods, but at the long interval between days of practice a drop occurs, which increases with the passing of the days. Thus, it is clear that the one-min. interval encouraged considerable growth that was permanent, that is, primary growth, but repetitions so close in time also encouraged a growth that was *temporary* in character which was increasing with practice.

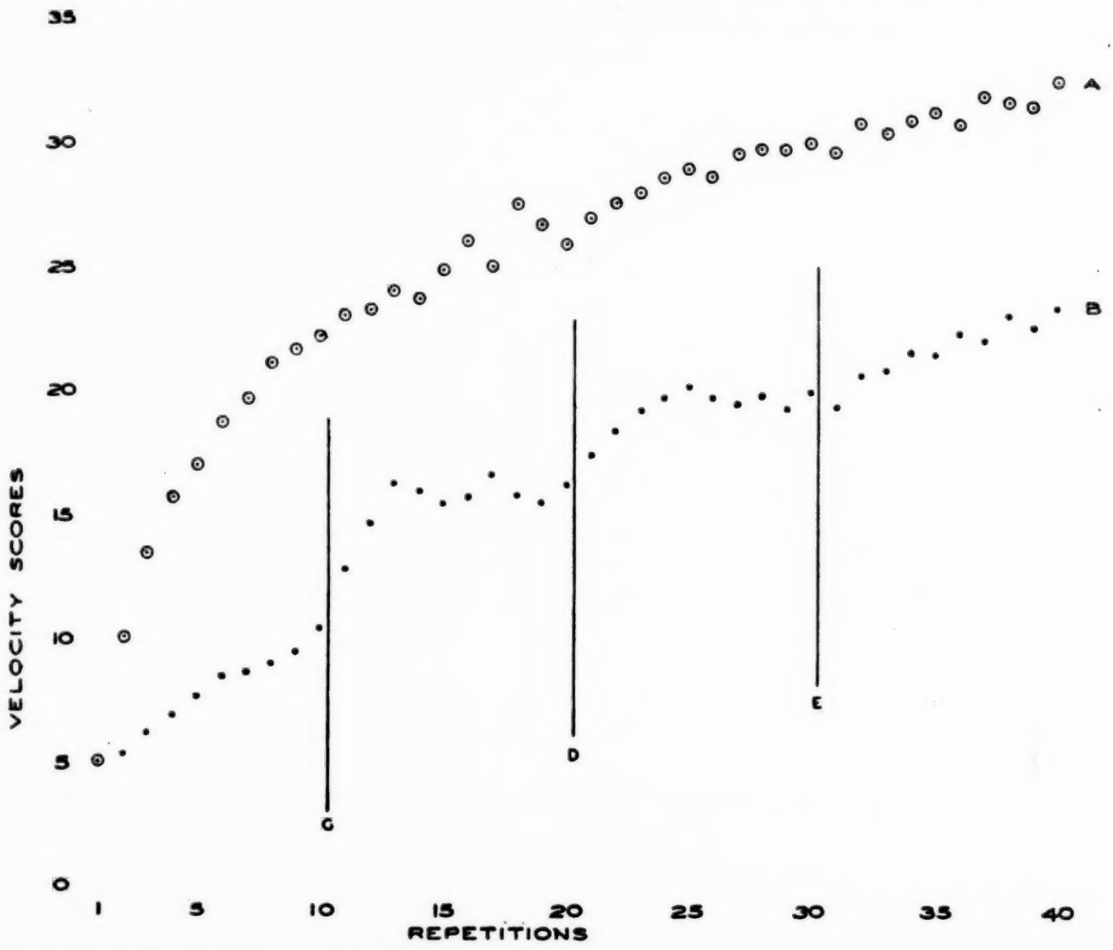



FIG. 3. Diagram showing the *A* and *B* groups. The *A* group makes one circuit per day. The *B* group makes 10 continuous circuits per day. The vertical lines *c*, *d* and *e* mark the 48-hour intervals in the *B* group. The equations are:

$$A \quad y = 5.82 + 16.89 \log x$$
$$B \quad y = 4.12\, x^{.48}$$

The primary growth came early and the secondary came later. This shows the early and late character of the two growths as they usually occur.

The outstanding power of the mirror vision coördination resides in its capacity to separate the growth coming from time intervals and that from continuous repetition. This is not possible in the case of the vast majority of learning studies, and even the recognition

of this issue is confused by the makers of these studies. This is the reason that the present author's work has met with some criticism from students of learning: their studies have apparently given different results. With the mirror vision coördination, a very simple activity may be repeated once or many times, and time may be admitted after each single repetition, or varied to suit one's purpose. *For the vast majority of learning activities this simple procedure is not possible.* Can Thorndike do this with his learning activities? Certainly not. Can Thurstone, Carr or the many students of learning do this? Certainly not. Thorndike vaguely saw this years ago (7), and established for the years a pathetic and costly error. He recommended that we keep practice periods constant, i.e., that each *repetition* be given a constant period of time, say five min. In mirror vision, the doing of an act *just once* is the unit. In Thorndike the doing of a most indefinite thing for a constant interval of time, say for five min., is the unit. Years ago the author found that this could be made clear to students by the following device: let *a* represent a doing of the act, and let a time interval be represented by a dash (–). Then the crowning achievement of the mirror vision coördination is that one can secure such a practice as a–a–a–a, etc.; while with Thorndike's type of activity, only this can be secured: aaaaa–aaaaaaa–aaaaaaaaa. The present work shows that the effect of even short time intervals plays an enormous role in the learning; while eliminating the time intervals, as Thorndike and most students of learning do, through their principle of holding practice time constant, changes the type of learning entirely. One can now see the rather doubtful procedure of Doré and Hilgard (2), and later of Bell (1), in attempting to refute our findings on the time interval by a Thorndike type of control.[2]

[2] The writer in 1938 pointed out this error in his reply to Doré and Hilgard (5), but these authors overlooked the criticism entirely. Later, Bell (1) in 1942, repeated the Doré and Hilgard experiment and the error of holding time constant, instead of the act, and even quoted our own criticism of Doré and Hilgard, without apparent insight into the meaning of the criticism.

In Doré and Hilgard's article, these authors resort to a type of criticism which requires some attention here. The writer had presented many learning curves of the type $y = a + b \log x$, all computed by the method of least squares. Later he sought to study the cumulative mean, the function formed by the encircled stations in Fig. 2, and found, empirically, that the same equation held. Now these authors point out that if we had only observed that, in the simple linear or straight line curve, the curve of the cumulative mean has a slope one half that of the linear equation (which is true), then, in our equation, which is in no sense linear, we should have for the cumulative mean half the slope of the first equation, that is, we should have $y = a + \frac{b}{2} \log x$. This step is absurd mathematically. The actual empirical results in such cases should have guided them, and in this case we cited one equation which shows the utter inadequacy of the half rule for the slope. In 4, p. 78, we presented a graph with the equation $y = 3.12 + 18.73 \log x$, and for the cumulative mean, $y = -3.68 + 17.69 \log x$; certainly a large departure for the half rule, 17.69 against 18.73. The actual fitting of the cumulative mean, p. 78, was one of the best the writer ever encountered in psychology; and that alone would have disproved the Doré and Hilgard observation.

## Two Groups Compared

Our 1939 experiment, planned originally to meet the criticism of Doré and Hilgard, consisted of two large learning groups. The first group, *A* (Fig. 3), was made up of 60 adult college students who made one circuit of the instrument per day for 40 days. The second group, *B*, was made up of 56 adult college students who made 10 circuits of the instrument in continuous series per day, with 48 hours between days of practice. Thus, this group had no time interval at all for 10 circuits of the instrument. Fortunately, the two groups are almost of equal initial ability, so we know at once that any differences are due to the conditions of practice. The *B* group, on the second circuit of the instrument, begins to show the absence of the time interval and at once indicates the much lowered rate of growth. We have no direct proof for it, but it is possible that by comparing the two groups one can see the appearance of an inhibition factor in the *B* group; at least it soon begins a low rate of growth which is not much changed for the four days of practice. In reality, this is the Thorndike type of control which always presents a lowered rate of growth resulting, as we believe, from over-stimulation.

It is without doubt the assumed role of fatigue which has kept psychology from true insight into the positive effect of time intervals. Thus, in an instance like this, it has usually been said that the *B* group is simply tired; this in spite of the fact that the appearance of a depression is coming too soon to justify the assumption of fatigue. When we cross the long 48 hour time intervals, we have perfect evidence of the fact that the *B* group is not recovering from a *temporary* depression; but is showing the appearance of primary growth arising from the time intervals. The writer's conception introduces a shock effect, arising immediately from stimulation, which may be overcome by primary growth at the time intervals, for the wholly invalid, yet time honored, notion of fatigue.

## Nature of the Shock Factor

The shock effect must be recognized as resulting in the largest degree from a new set of stimulations. The *S* first begins his reactions by employing movements or coördinations that would be appropriate for the direct vision situation. These reactions are not appropriate to the new situation, but they persist until the entire visual and motor situation becomes, to the *S*, a very intense affair. This is the shock factor. As a result of this, the direct vision coordinations are split up or disintegrated into their motor reciprocals, which provides a series of zig-zag movements. These enable the *S* to make a circuit of the instrument. The initial learning results

from these zig-zag movements, but not from a *selection* of the *good* movements which might give the *S* satisfaction, as in the Thorndike theory. There is nothing to select out, as each of the zig-zags is as good, or as bad, as the other. They all have a certain *rightness* about them, but it is difficult to see how this *rightness* could be specfically selected out, as in the case of a movement. Selection of the movements will not avail, but only a selection of the rightness.

We very early observed that the time interval enormously facilitated the learning in its early stages. But how could this be? The answer is simple and clear. The time interval has the effect of *contracting* the patterned reactions into a more condensed or stable whole. But *contracting* a series of reactions, run off in time, preserves the important *trend* or degree of *rightness* the individual reactions contained. This is true of any temporal series of reactions whatsoever, which eventually reaches a goal. Thus, the learning process at this stage is not actually a result of a specific selection of certain movements, but rather of a condensation or contraction of a sequence of movements, thus preserving a trend. [See Snoddy (**6**).]

Throughout this field one must come to see that shock arises from lack of adaptation. All stimulation is shocking, so long as there is adaptation to be made to it. We would not expect to see shock, for example, in the very easy direct vision reactions; but when time intervals are placed between repetitions in this direct vision learning, primary growth occurs. Thus, there was shock and there was adaptation. When we consider the visual perception of a house, as we shall do presently, there would seem to be no shock; but following a long continuous experience with the house, a three month's absence from it will produce a very marked contraction in the perceptual dimensions. Thus, there was shock in the perceptions and a time interval did stabilize the perceptual reactions.

## Study of the Two Groups

The two groups start out with approximately equal scores: the *A* group has an initial mean score of 5.24 mm. per sec. and the *B* group, 5.20. The very next circuit the *A* group stands at 10.69 and the *B* group at 5.84. Thus, the *B* group has lost five full units on the second circuit, after three min. of effort. The *B* group was 100 percent as efficient as *A* initially, and three min. later it was 55 percent as efficient. On the third circuit the *A* group has reached 14.20 and the *B* group 6.57. It can clearly be seen that a built up fatigue is not operating to produce these differences here. The evidence is now clear that the *B* group is suffering from the *shock of stimulation*, arising out of the learning situation. This is made even clearer by

a statistical picture (Table I). The median of the percent of *B* reaching *A* for the first day of 10 circuits is 0.45, for the second day, 1.5, for the third day, 3.6, and for the fourth day, 4.45. Thus, it is clear that the delayed growth in *B* is very gradually overcoming the lead of the *A* group. When the critical ratios (the difference of the means divided by the S.D. of the difference) are inspected, we see for the first group of 10 scores, corresponding to the first day for

TABLE I

THE PERCENT OF THE B GROUP REACHING THE MEAN OF A

| Circuit | % of *B* | Circuit | % of *B* |
|---|---|---|---|
| 1 | 50.0 | 21 | 0.0 |
| 2 | 6.4 | 22 | 5.3 |
| 3 | 2.0 | 23 | 5.3 |
| 4 | 0.6 | 24 | 5.3 |
| 5 | 0.3 | 25 | 7.1 |
| 6 | 1.0 | 26 | 3.7 |
| 7 | 0.0 | 27 | 1.8 |
| 8 | 0.0 | 28 | 1.8 |
| 9 | 0.0 | 29 | 3.6 |
| 10 | 0.0 | 30 | 1.8 |
| 11 | 0.0 | 31 | 1.8 |
| 12 | 1.8 | 32 | 1.8 |
| 13 | 5.0 | 33 | 1.8 |
| 14 | 6.5 | 34 | 1.8 |
| 15 | 1.4 | 35 | 7.1 |
| 16 | 0.2 | 36 | 7.1 |
| 17 | 2.6 | 37 | 8.3 |
| 18 | 1.0 | 38 | 3.6 |
| 19 | 1.5 | 39 | 7.1 |
| 20 | 1.2 | 40 | 5.3 |

group *B*, a critical ratio of 10.4, for the second day a ratio of 8.9, for the third day a ratio of 7.8, and for the fourth day a ratio of 8.7. Thus, in both of these statistical pictures there is clear evidence that the initially greatly depressed function in *B* is gradually recovering. Note that this is the reverse of what would happen if fatigue were operating. In fatigue, we would have at first very little depression which would increase with stimulation; but here we have at first very great depression which decreases as time passes.

When we study the equations of the two groups we see the same picture. The equation technique is in every way the ideal method for studying these two learning situations. In *A* the curve fits very well the algebraic equation $y = a + b \log x$ where $a$ and $b$ are constants. These two constants were found by the method of least squares, but very satisfactory results can be secured by the method of average equations presented in the author's study of the two growths (4, p. 37). In the *B* curve the equation is of quite a different order and can be fitted by $y = B x^n$ where $B$ and $n$ are constants.

This equation can be handled easily by taking logs of both sides, giving $\log y = \log B + n \log x$ which now is a simple, linear equation to be treated by the algebraic device shown above. The two equations with constants are:

$$(A)\ y = 5.82 + 16.89 \log x$$
$$(B)\ y = 4.12\, x^{.48}$$

These two equations have properties which are wholly different.[3] The initial constants 5.82 and 4.12 represent the first smoothed scores in both cases. The *A* curve begins immediately an enormous rate of growth which reduces slowly and eventually becomes very flat. The *B* curve proceeds from the start at a greatly reduced rate which is maintained, however, at a level eventually reaching that of the *A* curve. The lower curve will reach the upper somewhere between the 15th and 20th days of practice, or at about 150 repetitions. The two curves have become so flat that they may, for practical purposes, be considered identical from this point on. It must be kept in mind that the *B* curve was receiving very large primary growth increments at its 48-hour time intervals, but we have demonstrated completely the early non-fatigue character of the depression coming from the early continuous stimulations in the *B* group.[4]

## The Shock Factor and the Time Intervals

It has been pointed out that a depression factor results from the shock effects of stimulation. It is clear that the initial depression occurring at the first repetition is the same for all types of practice. The difference occurs later and the group with time intervals begins to show at once an increased *resistance* to the shock effect of stimulation, that is, this group is *shocked back* less. The shocking factor resides in the learning task itself. We do not consider that there is

[3] The difference between the two equations can best be seen by observing their occurrence in thermodynamics. If a gas be compressed rapidly, so that the heat formed is allowed to interfere, we have an example of the $y = Bx^n$ equation. But, if the gas be compressed slowly, so that the heat of the operation is allowed to escape, then the pressure relations will follow the equation $y = a + b \log x$.

[4] For those wishing to try the author's experiment, it is suggested that the profound effect of the time intervals compared with continuous stimulation, can best be shown if we employ the practice with intervals of one or two min., Fig. 2. Such a practice can be readily and easily carried out; while maintaining a practice of one circuit per day, for many days, is one of the most prodigious tasks to be undertaken. To do this latter, the writer required a crew of ten trained experimenters working over a year. But if one should use the one-min. interval, say, he will get a mixture of primary and secondary growth as shown in Fig. 2. The curve will go up in each day's practice, see Fig. 2, and then fall some at the long time interval. To get the real primary growth, the equation must be computed only for the initial circuits of each day. When the temporary growth is eliminated, the primary growth is found to vary positively with the length of the interval.

anything specifically different about our learning situation, aside from the fact that it is young, and on that account, highly shockable. It must be constantly kept in mind, also, that our device controls the learning situation so completely, that the stabilization due to the time interval is actually measured; while in many learning studies, the effect of the time interval is held off by the early unwitting *continuous* stimulation, which we have called the Thorndike effect.[5] Thus, these customary learning studies have only studied shock-curves, as our curve *B*, and as such, they are correct. The effect of the time interval is to *stabilize* the organism, that is, lead to such physiological changes as will increase its ability to endure that particular situation. This stabilization occurs, in our opinion, after any stimulation whatsoever, and provides, in perception, the stability or solidness of the physical universe. We have pointed out (4) that this growth, called stability, was in all probability the entropy of physical science, particularly as developed by Eddington (3). It will be recalled that Eddington (3, p. 104) saw entropy as a state midway between the physical and psychological processes. The role of time, in producing the stability of the physical universe, is the basis for the identity between the widely accepted physical process of entropy and primary growth in our function.

We have established the fact that primary growth is, in our opinion, due to the stability produced by the time intervals; but what produces secondary growth which results from bringing stimuli close together in time? This secondary growth is certainly *physiological summation*, which is well known in the study of reflexes. These two latter types of increment are identical as to temporary character and, we believe, as to cause.

The author proposed the primary and secondary growths without the slightest belief that they did the whole job in the field of learning. The evidence is clear that the author was right as to the two growths. There is every indication that something in the nature of *establishing associations*, that is conditioning, is necessary in psychology; but it is certain that primary and secondary growth factors will be associated with, and even condition, these other processes.

[5] The Thorndike effect is so marked that every laboratory should test it out. The mirror vision apparatus can be readily constructed and the test easily made. Practice should consist of two continuous circuits, then a time interval of two min., then two continuous circuits and a time interval. This can be repeated until 10 circuits and four time intervals have been secured. Now, compare this with another, unpracticed group making one circuit, followed by a time interval, until 10 circuits have been made. It will be seen that the one circuit and time interval is far superior; in other words, the presence of *two continuous circuits* is enough to shock back the primary growth very greatly. Thus, Thorndike's plan of doing an act over a period of time, then repeating it, actually secures for all time this shock effect. The error was unwitting, but its effect has blocked the conception of learning, and forever kept us from seeing the shock effect of stimulation and the stabilizing effect of the time interval.

## The Shock Factor and Perception

It has been pointed out that the shock factor appears in all perception, and in all learning involving perception. But why is this true? The writer has always been amazed by the fact that physiology has dealt almost exclusively with shock stimuli, and yet, psychology has never seemed to recognize this. Even the impulse passing along a nerve is essentially a negative state produced by shock; but how does it happen, that perception itself, has not revealed a shock factor? The answer here, in our view, is that the persistent failure of psychologists to recognize the part played by time, is responsible. One of the most widely observed phenomena has to do with the perceptual *contraction* of objects after the passing of a period of time, when the observer is away from the object. This is often referred to childhood, and many actually think it only occurs there, but this is not at all the case. The writer must beg the privilege of presenting a personal instance of this contraction occurring during mature years. Because of failing health, the writer left the Middlewest during the latter part of 1940 and moved to California to live. He was closely confined to his new home until the summer of 1943, when he returned to the Middlewest for a three month's visit. He was not aware that this would be a test of the phenomenon, and in fact, had not given it a thought. When he returned to Santa Monica, he was not aware of any perceptual change until his home was reached; and then it was clear that the building had *contracted.* But the greatest surprise occurred on going to the bedroom. The visual dimensions there had contracted at least 10 percent, it seemed. Now this is a well known phenomenon, and the instance is cited to show what can occur to one's perceptions, in mature years, after a three month's time interval. The contraction is in no sense reserved for childhood.

Why does the phenomenon occur, and why does the author make so much of it? The explanation is very simple. The perceptual apparatus is shocked in every act of perception, but we are aware of this only when the apparatus is stabilized after a considerable time interval; then, because of its greater stability, the apparatus is not shocked down so much, and the object of perception is sensed as smaller. This universally observed phenomenon is in complete accord with our explanation of the effect of the time interval, and shows conclusively, that primary growth continues indefinitely for perception and for learning.

## The Problem of Individual Differences in Learning

The writer has pointed out that individual differences in the mirror vision coördination present some unique trends (4). It was

shown that, in the time interval groups, individual differences are very much reduced, and that this constriction is much greater in early practice than in late. Also, in late continuous practice, individual differences are increased. The basis of the increase of individual differences is undoubtedly the pressure of stimulation; thus, it becomes clear that the large individual differences, secured by Thorndike and others, were actually produced by the pressure of stimulation maintained in their experiments. This pressure of stimulation has a definite value and significance in life, but its rôle in increasing individual differences must be understood. Thorndike's method of studying, unwittingly, only continuous stimulation, has given the whole problem of individual differences in psychology a false aspect. It becomes a necessary task for us to determine what our experimental devices contribute within *themselves*, before we can find the true nature of individual differences. This experiment of ours, then, has the effect of showing that such a task can never be fully accomplished, since the shock effect of stimulation must forever introduce a factor of uncertainty into individual differences in psychology, wholly comparable to the uncertainty principle in theoretical physics, which turns physics from an assumed absolute science into one wholly relative.

Since the reduction of individual differences is outstanding in the present research, it is chosen for presentation in this brief study. A large group of 70 university students were placed upon a practice of 20 repetitions per day, with time intervals of two minutes between successive repetitions; the practice covered a period of eight days. The score used was the mean velocity score for each day. The group was divided into fifths, in the order of the size of the individual mean scores for the first day of practice. From this point on, the individuals of each fifth remain constant for succeeding days. Then the mean of each group of 14 is divided by the mean of the entire group of 70 for each day. Thus, we get a picture of the behavior of each group in terms of the average for each day. Fig. 4 shows the results. The enormous reduction in individual differences is apparent at once. By the sixth day the middle three groups have had their average individual differences cancelled. From the slope of the lowest group it was suspected that this group would eventually reach the cancellation area. To get some light on this situation, this group was separated into an upper and lower half, with the results shown in the figure: the upper half enters the cancellation area but the lower half falls away after the first early constriction. Thus, our figure shows that in three and one half groups, or 70 percent, the mean of the small group equals the mean of the whole. Since the mean is approximately at the mid-point of the small groups, then the cancel-

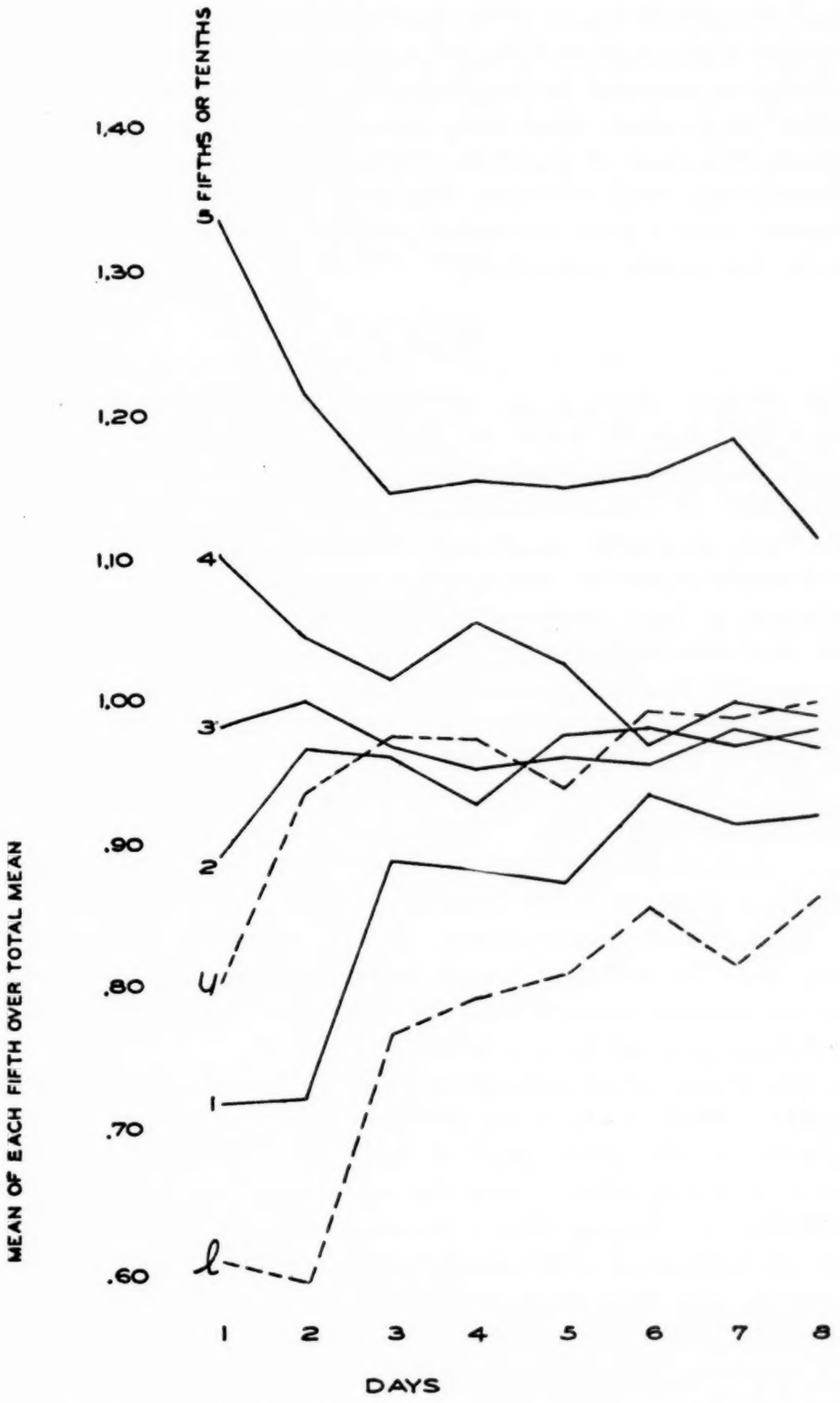



FIG. 4. Showing the reduction of individual differences in a group with long time intervals. The group is divided into fifths, according to size of mean score at the first day and, on later days, the subsequent mean of each original fifth is shown. The group marked (*u*) is the upper half of the lowest fifth, and (*l* )is the lower half. The enormous reduction of individual differences is apparent. Each day's practice consisted of 20 circuits, separated by two-min. intervals.

lation holds for about 35 percent of the whole group of Ss. Similar separation of the highest fifth into a high and low tenth did not change the picture as revealed by the high fifth. Thus, the cancellation area is limited to a group extending from the tenth to the eightieth percentiles in this type of practice, over a period of eight days, covering 160 repetitions, with two-min. intervals between repetitions. It did not appear that a more extended practice would increase the cancellation in this group appreciably.

## Summary

The present effort is an attempt to adapt a previous study (4) to meet the criticism made by Doré and Hilgard (2). The findings revealed the two growth processes previously announced, and served to show that the source of the Doré and Hilgard error lay in the unwitting step, originally made by Thorndike, in confining all learning to a consideration of shock-type situations, where the learning stimulation is kept *continuous*, thus greatly reducing the primary growth and overemphasizing the rôle of secondary growth.

It was shown that the learning situation, or even perception, has a shocking effect, and that a time interval following the stimulation stabilizes the organism, so that on later stimulation it is shocked down less by the learning situation; this constitutes primary growth in learning or perception. Secondary growth resulting from continuous stimulation is regarded as temporal summation. The establishing of the identity between primary growth in learning,and what takes place in perception, serves to make the principles of primary growth universal for all learning involving perception; but it does not confine learning to these principles. Some other principle, as conditioning, may be involved.

It was shown that individual differences are greatly reduced by interpolated time in the early part of learning, and are increased by stimulation in the latter part of learning. Since the stimulus has always a shocking effect upon the organism, and interpolated time a stabilizing or restoring effect, we are forever deprived of an absolute picture of individual differences. Our measuring devices must use stimulation, and time must be allowed to pass. The absolute frame for individual differences, implicit in the writings of Terman, Thorndike, Thurstone and others, must be forever subject to a principle of uncertainty.

In the present and previous studies the writer has shown what may emerge for learning if we are careful to avoid shocking down the primary growth arising from early time intervals. Throughout the past history of experiment in learning, the one obstacle to progress

has been the unfortunate bunching of practice effects. The full primary growth appears only when single, isolated repetitions of a task are secured. Obviously, many learning tasks can never give us insight into primary growth, since in these, a doing of the task can never be stated in other terms than 'doing a page of material' or 'working for a short interval of time.'

(Manuscript received February 20, 1945)

Bibliography

1. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, No. 243.
2. Doré, L. P., & Hilgard, E. R. Spaced practice as a test of Snoddy's two processes of mental growth. *J. exp. Psychol.*, 1928, **23**, 359–374.
3. Eddington, A. S. *The nature of the physical world.* New York: Macmillan, 1929, p. 361.
4. Snoddy, G. S. Evidence of two opposed processes in mental growth. Lancaster: Science Press, 1935, p. 103.
5. Snoddy, G. S. Reply to Doré and Hilgard. *J. exp. Psychol.*, 1938, **23**, 375–383.
6. Snoddy, G. S. An experimental analysis of a case of trial and error learning in the human subject. *Psychol. Monogr.*, 1920, **28**, No. 124.
7. Thorndike, E. L. *An introduction to the theory of mental and social measurements.* Teachers College, Columbia University, 1904.

# RELIABILITY OF BLINKING FREQUENCY EMPLOYED AS A MEASURE OF READABILITY*

BY MILES A. TINKER

*University of Minnesota*

Readability refers to the ease and efficiency with which material is read. It may be influenced by typographical factors such as size of type and by environmental factors such as illumination. During the past ten years various criteria of readability have been emphasized. Among these criteria is rate of involuntary blinking during the reading. Luckiesh and Moss (**2**, **3**, **4**) consider blink-rate to be a highly sensitive measure of readability and of ease of seeing during visual work.

The usefulness of any measuring device depends upon its reliability and validity. This experiment is concerned with the reliability of the blink-rate technique as a measure of readability. If valid conclusions are to be drawn from any set of data, the measurements involved must have an adequate degree of consistency, i.e., reliability. Our problem, therefore, is to determine under controlled experimental conditions the consistency of blink-rate from one reading period to another.

## Procedure

The experiment was carried out in a light laboratory with 10 foot-candles of well diffused illumination. Two studies were completed. In Study I, 74 university students served as readers, and in Study II, 64 students. The textual material was placed on an inclined stand at the normal reading distance from the *S*—about 14 in. In Study I the *S*s read in Dibble's *Strenuous Americans*. This is printed in 11 point type with 2 point leading in a 22 pica line width on rough surfaced paper. In Study II the reading material consisted of Winkler's *Morgan the Magnificent*. This is printed in 12 point type with 2 point leading in a 23 pica line width on rough surfaced paper.

The *E* was thoroughly trained in counting eye blinks prior to the experiment proper. Direct observation was employed. The observer was well off to the side, viewing the reader's profile against a black background. Counting was done by observing the movements of the upper eyelid. Partial as well as more complete closures were counted as blinks. A special timer was employed. Blinks were recorded for each of six successive five-min. periods of reading—30 min. in all.

On arriving at the laboratory, the *S* was adapted for 10 min. to the illumination used. He was told that the study was concerned with individual differences in reading. Only a few became aware that the blinks were being recorded. According to Luckiesh and Moss (**2**) such knowledge is not a deleterious experimental factor.

* Grateful acknowledgement is given to the Graduate School, University of Minnesota, for research grant to finance this study.

Each *S* was instructed to seat himself at a comfortable reading distance from the book, to start reading when the signal was given and to continue until time was called at the end of 30 min. He was told to read with comprehension and at his normal rate. The number of pages read was recorded. The *E* in Study II was different than in Study I.

## Results and Discussion

The basic data for Study I are given in Table I. The mean rate of blinking increases steadily through four periods (20 min.) and then levels out. The relative rate of blinking reveals this trend more clearly, taking the first period as 100 percent. Note that an increase of less than 5 blinks on the average yields a change of 25 percent.

TABLE I

Study I. Means, Relative Rates of Blinking, Standard Deviations and Coefficients of Variation for Blink-Rates. $N = 74$ Readers

| | Successive Five-Min. Periods of Reading | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Mean | 18.7 | 20.1 | 22.5 | 23.9 | 22.2 | 23.4 |
| Relative Rate | 100.0 | 107.5 | 120.4 | 127.9 | 118.8 | 125.2 |
| S. D. | 15.9 | 15.6 | 16.9 | 20.7 | 14.5 | 15.0 |
| V | 85.0 | 77.6 | 75.1 | 86.6 | 65.3 | 64.1 |

The standard deviations suggest that the group is markedly heterogeneous. Examination of the individual scores, however, reveals that the size of the standard deviations is strongly influenced by a few comparatively high scores. This tendency toward heterogeneity is more clearly shown by the coefficients of variation (V). They fluctuate between 64.1 and 86.6. Note that they are considerably less for the last two five-min. periods.

TABLE II

Study I. Reliability Correlations Between Blink-Rates During Reading Periods. $N = 74$ Readers

| Periods Compared | Coefficient |
|---|---|
| First 5 min. vs. Second 5 min. | .89 |
| Second 5 min. vs. Third 5 min. | .84 |
| Third 5 min. vs. Fourth 5 min. | .74 |
| Fourth 5 min. vs. Fifth 5 min. | .84 |
| Fifth 5 min. vs. Sixth 5 min. | .91 |
| First 5 min. vs. Sixth 5 min. | .56 |
| First 10 min. vs. Second 10 min. | .77 |
| Second 10 min. vs. Third 10 min. | .88 |
| First 10 min. vs. Third 10 min. | .66 |
| First 15 min. vs. Second 15 min. | .75 |

The reliability data (i.e., consistency of blinking from period to period) are given by the correlations in Table II. When the successive five-min. periods are adjacent to each other, as first vs. second or fifth vs. sixth, the reliability coefficients tend to be substantially high.

They fluctuate between .74 and .91. When one five-min. period is separated from the other by an interval of 20 min., however, the reliability coefficient is only .56, which is relatively low. It reveals a marked trend toward inconsistency in rate of blinking. That is, the reader who blinks fast in one period is apt to blink slower in the other, and vice versa. When blinking during adjacent 10-min. periods is correlated, the reliability coefficients are .77 and .88. But when there is 10 min. of reading between such periods, the coefficient drops to .66. This reliability coefficient is not high. For two successive 15-min. periods of reading the degree of consistency of response is shown by the coefficient of .75.

TABLE III

STUDY II. MEANS, RELATIVE RATES OF BLINKING, STANDARD DEVIATIONS AND COEFFICIENTS OF VARIATION FOR BLINK-RATES. $N = 64$ READERS

| | Successive Five-Min. Periods of Reading | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Mean | 25.2 | 29.6 | 31.5 | 31.3 | 32.5 | 34.3 |
| Relative Rate | 100.0 | 117.5 | 125.1 | 124.3 | 129.0 | 136.2 |
| S. D. | 14.7 | 18.1 | 19.9 | 19.7 | 20.4 | 24.9 |
| V | 58.3 | 61.2 | 63.2 | 62.9 | 62.8 | 72.6 |

The basic data for Study II are given in Table III. The mean blink rate increases fairly steadily for successive five-min. periods of reading. The relative rate shows that about 36 percent more blinks were made in the sixth period than in the first. The standard deviations and the coefficients of variation reveal heterogeneity of response in the group. As in Study I, this is somewhat influenced by a few very high blink-rates. Thus, in general, the trends are similar to those in Study I.

TABLE IV

STUDY II. RELIABILITY CORRELATIONS BETWEEN BLINK-RATES DURING READING PERIODS. $N = 64$ READERS

| Periods Compared | Coefficient |
|---|---|
| First 5 min. vs. Second 5 min. | .87 |
| Second 5 min. vs. Third 5 min. | .88 |
| Third 5 min. vs. Fourth 5 min. | .83 |
| Fourth 5 min. vs. Fifth 5 min. | .93 |
| Fifth 5 min. vs. Sixth 5 min. | .92 |
| First 5 min. vs. Sixth 5 min. | .49 |
| First 10 min. vs. Second 10 min. | .80 |
| Second 10 min. vs. Third 10 min. | .93 |
| First 10 min. vs. Third 10 min. | .62 |
| First 15 min. vs. Second 15 min. | .75 |

The reliability coefficients for blink-rates in Study II are shown in Table IV. They are very similar to those in Study I. For adjacent five-min. periods the coefficients range from .83 to .92.

Between the first and sixth periods the correlation is only .49, again revealing a tendency toward inconsistency in blink-rate when a 20-min. interval of reading intervenes. For successive 10-min. periods the consistency of blinking is high, but drops to .62 when 10 min. of reading intervenes. As in Study I, the consistency of blinking from one 15-min. period to a succeeding 15-min. period is substantial ($r = .75$).

In general, the data in both studies on reliability reveal the following: (1) Reliability, or consistency of response in rate of blinking is substantial or high only when adjacent successive periods of reading are involved. (2) When as much as 20 min. of reading intervenes between five-min. periods, the reliability drops to .49 or .56, revealing a tendency toward inconsistency in blink-rate. It is possible that this inconsistency would become more marked when longer intervals of reading intervene between initial and final five-min. periods in which blinks are counted. In many published studies blink-rates have been recorded for the first and last five min. during an hour of reading. (3) Consistency of blink-rates is approximately the same for adjacent 10-min. periods of reading as for adjacent five-min. periods. When an interval of reading intervenes, however, reliability of blink-rates is slightly higher for 10-min. than for five-min. periods.

TABLE V

STUDY II. MEDIANS AND SAMPLE INDIVIDUAL TRENDS IN BLINK-RATE

| | Blinks for Successive Five-Min. Periods of Reading | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Median..... | 20.0 | 23.0 | 26.0 | 25.5 | 27.0 | 27.0 |
| Subject A..... | 24 | 23 | 28 | 27 | 30 | 41 |
| Subject B..... | 18 | 17 | 17 | 19 | 19 | 18 |
| Subject C..... | 41 | 41 | 49 | 39 | 19 | 27 |
| Subject D..... | 46 | 69 | 74 | 58 | 54 | 50 |
| Subject E..... | 19 | 48 | 75 | 127 | 120 | 129 |
| Subject F..... | 33 | 34 | 35 | 33 | 30 | 65 |
| Subject G..... | 11 | 14 | 16 | 12 | 14 | 11 |
| Subject H..... | 33 | 37 | 49 | 28 | 23 | 24 |
| Subject I..... | 20 | 52 | 31 | 47 | 31 | 49 |
| Subject J..... | 17 | 17 | 24 | 26 | 34 | 34 |
| Subject K..... | 42 | 37 | 23 | 25 | 22 | 12 |

To throw further light on consistency of blink-rates the individual records of the readers in Study II were examined. Median blink-rates for the whole group and several typical sets of individual scores are listed in Table V. The trend of the medians is similar to that of means (Tables I and III). The sample individual scores reveal a variety of trends. Certain ones are similar in trend to the medians

(A, J), another shows little change (B), still another increases and then recedes to its original rate (G), and one fluctuates widely in succeeding periods (I). Subject K decreases rather consistently, while H rises and then decreases, F shows little change for five periods and then increases nearly 100 percent, and E increases by several hundred percent. It is obvious that the blink-rates of many individual readers are inconsistent from one five-min. reading period to another. Comparison of the blink-rate during the first five-min. period with that of the sixth period for all 64 readers reveals the same score for two readers, a lower score during sixth period than during first for 19 readers, and a higher score during sixth period than during first for 43 readers. In other words, some readers maintain the same blink-rate, some decrease their blink-rate, and others increase their blink-rate as reading continues. These results coincide with the findings of McFarland, Holway and Hurvich (4). "The blink-rate of some readers tends to increase as they continue to read. In others it decreases with an increase in reading time. Still others manifest no change at all."

The usefulness of a measuring instrument is dependent to a large degree upon its reliability. Where group comparisons are involved, the reliability coefficient of the measuring instrument should be .50 to .60 or above. If the measure is to be employed to differentiate among the scores made by individuals in a group, the coefficient must be .90 or above (1). For first vs. final five-min. periods in one-half hour of reading, the reliability coefficient for blink-rates barely reaches the minimum requirements for group comparisons ($r$ = .49 and .56). When the initial vs. the final 10 min. of a half hour of reading are compared the reliability is somewhat more satisfactory ($r$ = .62 and .66).

When adjacent (successive) periods of reading are involved, the reliability of blink-rates is relatively high for either five or 10-min. periods. Thus, the reliability requirements for group comparisons in this type of situation are adequate. In fact, the reliability is almost high enough for differentiation among individual scores.

The application of these findings to experimental studies in which blink-rates are recorded to measure ease of seeing is obvious. Many of the studies completed by Luckiesh and Moss are cited in their book (2). In these investigations where blink-rates are recorded for successive five-min. periods, such as in comparison of reading text set in light, medium and bold type-faces, the reliability of the blink-rate scores indicates adequate consistency of response for group comparisons of the data. However, in certain other experiments which involved comparison of blink-rates at the beginning and at the end of an hour of reading, the reliability is apt to be critically low, since

blink-rates barely reach the minimum reliability requirements at the end of 30 min. of reading. It is recommended that blink-rates be recorded for at least 10 min. when the reading periods concerned are separated by an interval of reading.

After prescribing conditions which will assure adequate reliability of blink-rates, it is necessary to establish the validity of this measure of readability if sound conclusions are to be drawn from the data. The validity of the blink-rate technique for use in studies on readability of print will be investigated in a study to be reported later.

## Summary and Conclusions

1. The purpose of this investigation is to determine the reliability or consistency of response for blink-rates as a measure of readability.

2. Two groups of readers were used, 74 in Study I and 64 in Study II.

3. The blink-rates were recorded during each of six successive reading periods of five min. each.

4. Trends in the two sets of data were approximately the same. The mean blink-rate tended to increase on successive periods of reading and the coefficients of variation were high.

5. The blink-rate consistency was fairly high from one five-min. reading period to the next period (median reliability coefficient = about .88). The trend was about the same for adjacent 10-min. periods of reading.

6. When an interval of 20 min. of reading intervenes, as between the first and sixth period, the reliability drops to about .50 for blink-rates during five min. of reading. For 10-min. periods of reading under similar conditions, the coefficient drops to about .65.

7. When individual data are considered it is found that, as reading goes on, the blink-rates of some readers increase, of others decrease, and of still others show no change. During successive five-min. periods of reading, individual blink-rates may oscillate markedly.

8. These data reveal highly satisfactory reliability where group comparisons of means are involved for blink-rates recorded during successive adjacent periods of reading. Where a period of reading intervenes between those in which the blinks are counted, however, the reliability of the blink-rates barely reaches the minimum requirements for comparison of group means. Since the reliability of the blink-rates under these conditions increases somewhat (to about .65) for 10-min. periods, it seems best to record blinks for 10 rather than for five min. when a period of reading intervenes between the periods compared.

9. Published studies which have compared blink-rates for successive five-min. periods which are adjacent to each other may be

considered to have attained adequate reliabilities in their measurements. In those studies where 50 min. of reading intervened between the five-min. periods compared, the reliability apparently fluctuates close to the level of undesirable inconsistency.

10. The next step is to determine the validity of blink-rates as a measure of readability.

(Manuscript received April 18, 1945)

## References

1. Garrett, H. E. *Statistics in psychology and education.* New York: Longmans, Green and Co., 1937, pp. 493.
2. Luckiesh, M., & Moss, F. K. *Reading as a visual task.* New York: D. Van Nostrand Co., 1942, pp. 428.
3. Luckiesh, M., & Moss, F. K. Criteria of readability. *J. exp. Psychol.*, 1940, **27**, 256–270.
4. Luckiesh, M., & Moss, F. K. Frequency of blinking as a clinical criterion of ease of seeing. *Amer. J. Ophthal.*, 1939, **22**, 616–621.
5. McFarland, R. A., Holway, A. H., & Hurvich, L. M. *Studies of visual fatigue.* Boston (Soldiers Field): Graduate School of Business Administration, Harvard University, 1942, pp. 255.

# ARTIFACTS IN THE INVESTIGATION OF SENSITIVITY TO VIBRATION[1]

BY LOUIS D. GOODFELLOW

*Northwestern University*

In the past 50 years, the study of the perception of mechanical vibration (**1, 3**) has received the attention of 65 authors, who have contributed 170 papers to the professional literature on the various aspects of the subject. A number of serious discrepancies are found in the actual data presented, and especially in the interpretations given them. Since most of the authors mentioned are reliable research men who would not be apt to report data carelessly collected or naively interpreted, it seems likely that there was something basically wrong with the experimental techniques used. Rather than take sides in these controversies, the present author has chosen to examine these techniques in an attempt to find the source of the discrepancies.

In the literature we find that most experimentalists have reported thresholds in terms of the amplitude of the vibration, the implied assumption being that the actual stimulus is caused by a depression of the skin. Other experimenters have preferred a measurement of the energy dissipated, and have reported their threshold values in decibels below some arbitrary level. Since the energy of a vibrating system is a function of a number of factors in addition to amplitude, it becomes evident that the two measures of intensity are not equivalent. A laboratory demonstration of this proposition may be observed in the stimulus-value of two vibrators, one of which has a considerably larger contact area than the other. (The larger contact area will necessitate increased mass in order to maintain a reasonable impedance match.) The thresholds, in terms of energy consumption for these two vibrators will be roughly the same—which means that the amplitudes must differ widely because the larger vibrator consumes considerably more energy per unit amplitude of vibration. Actual measurements of the amplitude of vibration made while in

[1] These studies were initiated in 1936 in the Physiology Department of the Northwestern University Medical School. They were made possible by two grants from the Committee on Scientific Research of the American Medical Association. The original data were equivocal and difficult to interpret without further experimentation. This additional research was made possible in 1940 by a grant from the Audio-Visual Research Foundation. The author expresses his deep appreciation to Dr. A. C. Ivy for his generous coöperation in this work.

contact with the finger confirm this statement. We must admit considerable bewilderment on first observing the distinctly different stimulus-values of two vibrations of identical amplitude. Attempts to observe this phenomenon with various designs of vibrators show that only the large massive vibrators gave thresholds which appeared to correlate with the energy consumption. In fact, most of the lighter and more common designs produced thresholds which appeared to correlate with the amplitude of vibration.

The first important consideration in explaining these observations is the frequently-overlooked fact that the stimulus applied to the skin is not necessarily the equivalent of that reaching the end-organ. The difference is due, we believe, to the transmission of mechanical energy from the vibrating body to the end-organ. This transmission will depend upon the proper matching of mechanical impedances, which in turn depends upon several considerations, namely, the area of the contactor, the elasticity of the tissues involved, and the design of the vibrator minimizing the increment of reflected impedance. It is common knowledge that a small pin vibrating up and down in a pool of water will transmit very little energy to the water. If, however, a plunger be added to the end of the pin, a great deal of energy can be transferred to the water.

The small contactor, shall we say, is analogous to the pin without the plunger, and, in this case, very little energy is transmitted to the tissues. Since the distortion of tissue is limited largely to the point of contact, the actual stimulus is the extent to which a receptor in the skin is displaced. Consequently, with this type of vibrator, it seems reasonable to expect that the amplitude would be more closely correlated with sensitivity. On the other hand, the larger type of vibrator (analogous to the pin with the plunger added) will transmit a large proportion of its energy to the tissues, and the superficial receptors will ride up and down with the large mass of tissue, thereby remaining practically undistorted and consequently unstimulated. The end-organs stimulated are probably those farther removed from from the source of vibration, such as those imbedded in the deeper tissues.

The extent to which these end-organs are distorted can hardly be dependent upon the amplitude of the vibrating contactor, since different sized contactors, properly coupled to the skin, transmit different amounts of energy per unit amplitude of vibration, and the amplitude of vibration in the tissue at any particular distance from the source is a function of the energy being dissipated.

The mechanical impedance looking 'into the skin' is largely resistive (at least for the finger-tip), and therefore considerable dissipation of energy results. It is unfortunate for the study of

sensitivity to vibration that no efficient transmission channel (such as the middle-ear provides for hearing) exists between the end-organ and the external stimulating medium.

If our reasoning that different types of vibrators stimulate different end-organs is correct, then the results obtained are interpretable only with reference to the particular apparatus used. This is our explanation of the conflicting reports found in the literature.

## Procedure

Let us turn then to the original experiments[2] which led to the speculation in which we have ust engaged. The purpose of the study was:

1. To determine the sensitivity of the sole of the dog's foot to mechanical vibrations of various frequencies and thus afford a comparison to human beings.
2. To establish a frequency-intensity function for the dog.
3. To determine the stability of threshold values from day to day.
4. To determine the influence, if any, of hearing on thresholds.

For this research, unselected female dogs of medium size were trained to lie quietly on a table without any hampering apparatus.[3] After they had learned to lie quietly for half-hour periods, a vibrator was applied to the sole of one foot, and a conditioned withdrawal reflex of the homolateral limb was set up.

After a conditioned reflex was well established, the intensity was attenuated in steps of decibels until the animal failed to respond—appropriate reinforcements being made to prevent the extinction of the reflex. The intensity yielding responses 75 percent of the time was taken as the threshold value.

The equipment consisted of a variable pitch audio-oscillator with appropriate meter and attenuators for regulating the intensity of the vibratory stimulus. The unconditioned tactile stimulus was applied to the dog's feet by means of an electric magnet. Both conditioned and unconditioned stimuli were regulated by a pendulum to provide a proper temporal relationship. A check on the actual vibration of the contactor was made by a technique similar to that described by Geldard and Gilmer (2) in which the vibrations are picked up by a crystal vibration detector and analyzed on the cathode-ray tube.

Because many different types of vibrators were being used in research in this field, it seemed desirable to employ two radically different types[4] for these studies. One was a small Utah magnetic speaker motor to which a pin (area of one square mm.) was attached to serve as the contactor. The other instrument consisted of two large Utah theatre-speaker motors. They propelled a metal cone mounted on a suspension of high-grade steel having a natural period of 6050 c.p.s. and to which a contactor (area 250 mm.$^2$) was attached.

The dogs represented by large letters were tested with the larger vibrator and those animals designated by small letters with the small instrument. All other conditions were identical for all animals. Three medical students served as laboratory assistants, one working with animals *A* and *a*, the second with animals *B* and *b*, and the third with animals *C* and *c*.

## Results and Discussion

Table I shows that animal sensitivity was practically that of the human *S*,[5] the slight superiority of the animal being explainable

[3] Part of this research was reported to the Midwestern Psychological Association in May, 1939, under the title, "A Preliminary Report on the Sensitivity of the Dog's Foot to Mechanical Vibration."

[3] The author considers any restraining mechanisms a hazard to a psychological experiment.

[4] Both of these vibrators were designed and built for this study by A. W. Krause, then of the Department of Physics, Northwestern University.

[5] The thresholds for human subjects in terms of amplitude range from .01 to 10 microns depending upon a number of factors.

by the different definitions of threshold implied by the different method used for finding it.

Turning to the second question, namely, the frequency-intensity function, we notice from Table I that it is essentially the same as that reported for human observers. The failures of animals *a*, *b*, and *c* to respond to conditioning at frequencies above 2048 c.p.c. can undoubtedly be explained by the facts that the area stimulated was comparatively small and the amplitude of the vibration too restricted.

TABLE I

THE SENSITIVITY OF THE DOG'S FOOT TO VIBRATIONS IN DECIBELS BELOW THE HUMAN THRESHOLD AS DETERMINED BY THIS LABORATORY

| | *A* | *B* | *C* | *a* | *b* | *c* |
|---|---|---|---|---|---|---|
| 32 | 4 | 1 | 3 | 1 | 0 | 1 |
| 64 | 6 | 3 | 4 | 2 | 1 | 1 |
| 128 | 4 | 3 | 4 | 2 | 1 | 1 |
| 256 | 3 | 2 | 2 | 1 | 0 | 1 |
| 512 | 3 | 3 | 2 | 1 | 0 | 0 |
| 1024 | 2 | 3 | 1 | 0 | 0 | −1 |
| 2048 | 4 | 2 | 4 | −1 | −2 | −1 |
| 4096 | 6 | 4 | 6 | — | — | — |
| 8092 | 8 | 6 | 6 | — | — | — |

The third question, the stability of thresholds, was planned to throw light on the problem of the great instability found among human *Ss*. We hoped to answer the question as to whether the variation from day to day was due to psychological causes which could be controlled in animal studies or to factors inherent in this form of sensitivity. At first the data appeared disappointingly equivocal, but more careful examination showed that the animals using the large vibrator were remarkably stable, but those using the small instrument resembled human *Ss* in their instability. We were inclined to attribute this to individual differences among the animals, but ultimately decided to test the proposition by trading vibrators. The results were exactly reversed and stability was definitely related to type of vibrator. The explanation, although puzzling at the time, seems clear in the light of subsequent findings. On the assumption that the small vibrator (one mm.$^2$) stimulates primarily touch spots, it is assumed that we did not stimulate exactly the same spot on different days. In the case of the large unit (250 mm.$^2$), a slight variation in position would be negligible, since the primary stimulation is in the deeper tissues. It was these discrepancies between different vibrators which started us on the research that ultimately led to the considerations set forth in the beginning of this paper.

Now let us turn to our fourth objective, namely, determining the influence of hearing on this form of sensitivity. For many years in

working with human *Ss* the author has been haunted by the possibility that subliminal auditory stimuli either from aerial radiation or conduction through the body tissues to the cochlea might be operating. It was planned to settle this question surgically by destroying the cochlea and noting the effect in vibratory thresholds.

After the destruction of one cochlea, the animal was allowed a week's rest to recover from the ataxia. After a week's training to keep the conditioned reflexes intact, the opposite cochlea was destroyed.

After the animals were deafened, none showed any loss in sensitivity to vibration. In fact, animals *a*, *b*, and *c* showed slight but statistically reliable gains. It is interesting that the three animals using the large vibrator showed no change, but the three using the small vibrator showed gains in sensitivity after being deafened. These data might be taken as a definite negative answer to our fourth objective, relating to the effect of hearing on sensitivity after being deafened. However, any such conclusion is premature in the light of subsequent research.

TABLE II

EFFECT OF COCHLEA DESTRUCTION ON SENSITIVITY TO A VIBRATION OF 1000 C.P.S.

| Animal | Change in Decibels | Sigma of Difference | C.R. |
|---|---|---|---|
| *A* | 1 | 2.1 | .5 |
| *B* | 2 | 2.0 | 1.0 |
| *C* | −1 | 2.1 | .5 |
| *a* | 8 | 2.2 | 3.6 |
| *b* | 6 | 2.0 | 3.0 |
| *c* | 6 | 2.1 | 2.9 |

It is to be regretted that we were never able to follow up the point with additional experiments to explain the phenomenon, but certain speculation based on an incompleted study of inter-sensory facilitation and inhibition may be permitted at this point. It has been shown (**4**) that stimulation of one sensory modality alters the sensitivity of another. The precise inter-sensory relationships are not yet clear, except in a few cases. One of these cases in which our data are clear shows that auditory stimuli between plus two and plus six decibels facilitate vibratory sensitivity to the same pitch, but have inhibitory effects when the pitches of the auditory and vibratory stimuli differ. If the findings in our study of inter-sensory effects in human *Ss* hold true for dogs, our data are explainable as follows. Room noises, particularly the slight 120 cycle purr of our equipment, served as an antagonistic stimulus and raised the animals' threshold to vibratory stimuli. After destruction of the chochlea, this inhibi-

tory effect was removed and as a result thresholds dropped as is shown in Table II. The next fact to be explained is that the increases in sensitivity were much greater in animals *a*, *b*, and *c*. One way of accounting for this fact would be to assume that for the animals using the large vibrator, some sound reached the cochlea, and because it was the same pitch as the vibratory stimuli, had a facilitating effect This facilitation effect would tend to cancel the inhibitory effect in the normal animal. After the cochlear destruction both effects would be removed, resulting in little change in vibratory thresholds.

The possibility of conduction of vibrations through the body tissues to the cochlea is reasonable, since it is known that vibrations may be transmitted great distances in a human *S*. This possibility is much greater for animals using the large vibrator where a great mass of tissue was set in motion and would account for the failure in animals using the point stimulation where very little energy was transferred to the tissues.

To return to the fourth question which the research was planned to answer, i.e., what is the effect of hearing on sensitivity to vibrations, we must say that our data are equivocal on the point. Although no animal showed a loss of sensitivity after cochlea destruction, we have shown that there still is a possibility that there was some conduction to the cochlea, the effects of which were canceled by another factor—room noise. Until the experiment can be repeated in a soundproof room so as to eliminate the possibility of cancellation of these effects, we decline to commit ourselves to an explanation.

## Summary

Because of discrepancies appearing in the literature, these studies were undertaken to check certain data on the perception of mechanical vibrations, using animal *S*s so as to make possible more rigid controls.

We found dogs to be very serviceable subjects for research in this field, and developed a satisfactory technique. The dog possesses approximately the same range and degree of sensitivity as the human *S*. It yields a more stable threshold, our inference being that psychological factors, the prime causes of unstable results, are controlled. Our data on the possibility of subliminal auditory clues affecting vibratory thresholds are inconclusive, but these experiments led to an important technical discovery, viz., that results may at times be a function of instrumentation. The important consideration presumably is the amount of energy from the contactor which actually reaches the end-organ and which depends upon the proper matching of mechanical impedances. Failure to consider this problem in

instrumentation, we believe, accounts for many of the discrepancies found in the professional literature.

(Manuscript received March 2, 1945)

## References

1. Geldard, F. A. The perception of mechanical vibration. *J. gen. Psychol.*, 1940, **22**, 243–307.
2. Geldard, F. A., & Gilmer, B. von H. A method for investigating the sensitivity of the skin to mechanical vibration. *J. Psychol.*, 1934, **11**, 301–310.
3. Goodfellow, L. D. Vibratory sensitivity. *Psychol. Bull.*, 1934, **31**, 560–571.
4. Goodfellow, L. D. Significant incidental factors in the measurement of auditory sensitivity. *J. gen. Psychol.*, 1945. (In press)